TU Delft

Delft University of Technology

The effects of perceived value and stakeholder satisfaction on software project impact

Huijgens, Hennie; van Deursen, Arie; van Solingen, Rini

DOI
10.1016/j.infsof.2017.04.008

Publication date
2017

Document Version
Submitted manuscript

Published in
Information and Software Technology

Citation (APA)
Huijgens, H., van Deursen, A., & van Solingen, R. (2017). The effects of perceived value and stakeholder satisfaction on software project impact. Information and Software Technology, 89(September), 19-36. https://doi.org/10.1016/j.infsof.2017.04.008

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Delft University of Technology
Software Engineering Research Group
Technical Report Series

The Effects of Perceived Value and Stakeholder Satisfaction on Software Project Impact

Hennie Huijgens, Arie van Deursen and Rini van Solingen

Report TUD-SERG-2017-001

TU Delft

SERG

TUD-SERG-2017-001

Published, produced and distributed by:

Software Engineering Research Group
Department of Software Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology
Mekelweg 4
2628 CD Delft
The Netherlands

ISSN 1872-5392

Software Engineering Research Group Technical Reports:
http://www.se.ewi.tudelft.nl/techreports/

For more information about the Software Engineering Research Group:
http://www.se.ewi.tudelft.nl/

Note: Accepted for publication in *Information and Software Technology*, 2017. DOI https://doi.org/10.1016/j.infsof.2017.04.008.

© 2017 This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/

The Effects of Perceived Value and Stakeholder Satisfaction on Software Project Impact

Hennie Huijgens[a, 1], Arie van Deursen[b], Rini van Solingen[c]

[a] *Delft University of Technology and Goverdson, Mekelweg 4, 2628 CD Delft, The Netherlands*
[b] *Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands*
[c] *Delft University of Technology and Prowareness, Mekelweg 4, 2628 CD Delft, The Netherlands*

ARTICLE INFO
Keywords:
Software Economics
Evidence-Based Software Engineering
Perceived Value
Stakeholder Satisfaction
Cost Duration Index

ABSTRACT
Context: In this paper we present a multiple case study on the insights of software organizations into stakeholder satisfaction and (perceived) value of their software projects. Our study is based on the notion that quantifying and qualifying project size, cost, duration, defects, and estimation accuracy needs to be done in relation with stakeholder satisfaction and perceived value.
Objectives: We contrast project metrics such as cost, duration, number of defects and estimation accuracy with stakeholder satisfaction and perceived value.
Method: In order to find out whether our approach is practically feasible in an industrial setting, we performed two case studies; one in a Belgian telecom company and the other in a Dutch software company.
Results: In this study we evaluate 22 software projects that were delivered during one release in the Belgian telecom company, and 4 additional large software releases (representing an extension of 174% in project size) that were delivered in a Dutch software company. Eighty-three (83) key stakeholders of two companies provide stakeholder satisfaction and perceived value measurements in 133 completed surveys.
Conclusions: We conclude that a focus on shortening overall project duration, and improving communication and team collaboration on intermediate progress is likely to have a positive impact on stakeholder satisfaction and perceived value. Our study does not provide any evidence that steering on costs helped to improve these. As an answer to our research question - how do stakeholder satisfaction and perceived value relate to cost, duration, defects, size and estimation accuracy of software projects? – we found five take-away-messages.

1. INTRODUCTION

An often cited result of the 1994 Standish CHAOS research [1] is that 70% of all software projects are problematic. Standish defines these as so-called 'challenged projects', meaning they were not delivered on time, within cost, and with all specified functionality [2].

This is in a certain way along the lines of what we found when studying a series of 22 finalized software projects in a Belgian telecom company. We found that the average cost overrun was 28% (ranging from -41% to 248%), and that the average duration overrun was 70% (ranging from 9% to 168%). There was only one single project that performed within a 10% cost and duration overrun boundary. As such, these projects were challenged if we adopt the way Standish defines success and failure; being the extent in which a project conforms to its original plan.

However, did all the other 21 projects fail? Is it fair to say that a project with cost overrun is a failure? Is it reasonable to say that a project that performed completely according to plan, but delivered software that no one uses, is a success?

1.1 Problem Statement

Supported by many critical reviews of the Standish criteria [2] [3] [4], we define success and failure in this paper from a different angle, trying to include the balance between *value* and *cost* into the equation. In previous research we defined success and failure in terms of cost, duration and number of defects of

[1] Corresponding Author
E-mail Address: h.k.m.huijgens@tudelft.nl

a software project [5] [6] [7]. Looking at the outcomes of this we consider that a project that is late and over budget – and thus in terms of our study bad practice, or in other words unsuccessful – yet returns high value according to its stakeholders, may still be called successful, because of the fact that it delivers high value.

By analyzing project metrics such as *cost*, *duration*, *defects*, and *size* of the projects in connection with *stakeholder satisfaction*, *perceived value* and *quality of estimations*, we show that stakeholders define success and failure of a project different from solely measuring cost and duration overrun. Especially in domains where value is more important than predictability, e.g. agile ways of working, a limited view on conformance to planning, seems illogical.

Due to the fact that measuring the real – delivered – value of software deliveries is difficult, we focus in this paper specifically on *perceived* value. The underlying idea is that, since finding evidence in the bottom-line financial administration is hard, if not impossible, the best we can do is involve stakeholders for a qualitative indication of value. However, as this is strongly dependent on the individual and the contextual setting (what is valuable in one setting might not be valuable in another, or what one stakeholder considers to be of no value can be of high value to another stakeholder), we use the term *Perceived Value*. We understand that this is a way to measure value that is limited in its external and construct validity. However, this approach may help in finding early ways of indicating value [8].

In this paper, we analyze a set of projects conducted at a Belgian telecom company (referred to in this paper as BELTEL) and a Dutch software company (referred at in this paper as DUTCHCO) that provides billing software products and services (also largely to the telecom domain). We propose the following research question:

How do stakeholder satisfaction and perceived value relate to cost, duration, defects, size and estimation accuracy of software projects?

In answering this question, we make the following contributions:

1. We propose a light-weight value measurement technique based on post-release interviews.
2. We provide data on 26 industrial projects for which 83 key stakeholders provide *stakeholder satisfaction* and *perceived value* measurements in 133 completed surveys.
3. We contrast the resulting *perceived value* and *stakeholder satisfaction* statements with collected data on *costs*, *duration*, *defects*, *size* and *estimation accuracy* and look for links between them.

This paper is an extended journal version of an earlier published paper at the 20th International Conference on Evaluation and Assessment in Software Engineering (EASE 2016) **[6]**. Compared to the original paper the new contributions can be summarized as follows:

- We replicated the research performed in our original study in another company: DUTCHCO, a Dutch software company, specialized in delivering billing solutions to European telecom operators.
- Within DUTCHCO, we examined four (4) large software releases, representing an extension of 174% in project size. We collected detailed size, cost, time, and defects data from all releases. We performed electronic surveys on stakeholder satisfaction and perceived value among thirty (30) stakeholders within the DUTCHCO organization and INVEND, its provider of India-based development teams.

The remainder of this paper is structured as follows. In Section 2 related work and the background of the model that we use for analysis purposes are described. Section 3 outlines the research design. The results of the study are described in Section 4. We discuss the results in Section 5, and finally, in Section 6 we make conclusions and outline future work.

2. BACKGROUND AND RELATED WORK

Many studies include critical reviews of the Standish Chaos Report [2] [3] [4] [9] [10] [11] [12] [13]. The Standish Group reported in their 1994 CHAOS report that the average cost overrun of software projects was as high as 189%. Jørgensen and Moløkken-Østvold [2] conclude that this figure is probably much too high to represent typical software projects in the 1990s and that a continued use of that figure as a reference point for estimation accuracy may lead to poor decision making and hinder progress in estimation practices [2]. Glass [3] states that objective research study findings do not, in general, support those Standish conclusions [3].

Where in our research we measure value as perceived by stakeholders on four business related subjects, many different measures are used to identify value, and a clear and uniform definition is no question yet. Pekki [14] defines stakeholder value as the "usefulness of offering SPI to its key beneficiaries, so they are fully involved into SPI activities which increases the success of those activities". Beck [15] indicates that value is about money and time, by saying we "need to make our software economically more valuable by spending money more slowly, earning revenue more quickly and increasing the probably productive lifespan of our project". Dingsøyr and Lassenius [16] answer the question "What is value"? by saying that "the improvement trends are not specific on how they define value". They come up with the argument that," proponents of agile development would argue that a development team needs to learn what external stakeholders value during a development project". In a way this matches our idea that besides internal stakeholders, especially external stakeholders should be involved in the value discussion.

Atkinson [17] argues that besides time, cost and quality, often referred at as *the iron triangle*, also stakeholder benefits should be taken into the equation. Besides that, he mentions the effect that quality is "an emergent property of people's different attitudes and beliefs, which often change over the development life-cycle of a project".

Estimate the value of software is probably as challenging as predicting the cost of software [18]. Strand and Karlsen [19] suggested to estimate value in the form of "benefit points", as a kind of equivalent to story points. Cheng et al. [20] describe an architecture-based approach to discover value of software

engineering by using big data techniques. Although quite some research has been performed in the area of value estimation [21] [22] [23], and success criteria for software projects [24] [25], most of these approaches seem poorly adopted in industrial software project management settings. A good sign however, is that an increased focus on value in improvement is seen in software development, mainly driven by agile development approaches [16].

Jørgensen [26] performed a survey among software professionals in Norway on the characteristics of projects with success in delivering client benefits. He mentions that a focus on client benefits as a success criterion is particularly important, because only weak correlations are found on other dimensions, such as "being on time" and "being on budget". Besides that, he mentions that the traditional success factor "having the specified functionality" may even be in conflict with success in delivering client benefits.

3. RESEARCH DESIGN

The goal of this study is to understand the underlying reasons of stakeholder satisfaction and value of software projects. To achieve this, we contrast project metrics such as *cost*, *duration*, *number of defects* and *estimation accuracy* with *stakeholder satisfaction* and *perceived value*. We argue this will help to better understand the backgrounds of software projects as a guide for building future software portfolios.

As explained in the introduction, the Standish criteria [1] states that success and failure are related to the quality of project estimates. In order to explore alternatives, we test for association between paired samples, using Pearson's product moment correlation coefficient and resulting p-values in case our data is normally distributed or Spearman Rank Correlation when the data is not normally distributed. To mitigate the risk that we find coincidental correlations we perform an exploratory study that confronts correlated metrics with findings from qualitative results from analysis of the free format text from the surveys.

We performed a multiple case study in two different companies: BELTEL, a Belgian telecom company, and DUTCHCO, a Dutch software company that delivers billing solutions to European telecom operators. In the following two paragraphs we describe the industrial context of how both companies are included in our research.

3.1 BELTEL

BELTEL is a Belgian telecom company that can be characterized as a typical mid-sized information-intensive company with a mature software delivery organization that offers a mix of delivery approaches, ranging from plan-driven to agile (Scrum) [27]. For the majority of its software development activities BELTEL has a strategic, long-term contract with one large Indian supplier, referred to in this paper as INDSUP. Projects relate to different business domains (e.g. Internet, Mobile Apps, Data warehouse, Billing, Customer Relationship Management).

During the past three years, BELTEL has adopted a metrics program to collect data on size, cost, duration, the number of defects, and the estimation accuracy of finalized software projects. This data has been used to analyze project performance at BELTEL, to benchmark project performance, and to continuously improve the software delivery process within BELTEL. In October 2015, BELTEL changed its strategic focus from cost-based (steering on efficiency and operational excellence) to value maximization and shortening time-to-market. To facilitate this, BELTEL has collected additional data, addressing business value and customer satisfaction.

In the present paper, we compare these with the data on costs and duration that were also collected, in order to better understand the relationships between various project success indicators. Development projects at BELTEL are conducted independently, yet are grouped for deployment into so-called releases. Once a project passes its systemtest it is promoted to a release, which typically contains multiple projects. Releases are further tested and deployed as a whole. Within BELTEL eight subsequent releases are performed each year. In this paper, we study data from 22 projects coming from four different releases.

3.2 DUTCHCO

DUTCHCO is a Netherlands based software company that offers billing solutions to a large variety of European telecom companies. Within this market DUTCHCO is a European market leader.

Unlike BELTEL, DUTCHCO does not structure its work into projects. All software development activities are organized into four large market releases each year. Driven by the desire of its customers to limit the number of deployments, DUTCHCO implements only four market releases a year. As a result, these four releases are usually quite large in size. Where BELTEL thus implements eight releases a year, each of which consist of a large number of small and medium-sized projects, DUTCHCO performs only four large releases, which are composed of many small user stories.

To build and test its software, DUTCHCO makes use of several development teams in India [28]. These teams are supplied and supported by INVEND, a Dutch consultancy company, specializing in agile software delivery. Activities such as preparation of releases, design, quality assurance, and overall management are performed by members of an onsite, Netherlands based team of DUTCHCO itself.

Based on the results of previous research within the organization, DUTCHCO pays considerable attention to communication between the different members of a development team. There is a virtual contact window that is constantly open to allow team members in different locations to contact colleagues, and substantial effort is put into reciprocal visits to the team sites.

All teams within DUTCHCO – including the development teams of INVEND in India – work according to the Scrum approach [27]. An enterprise backlog and sprint backlogs are maintained in Jira, bi-weekly sprints are performed, results are demonstrated to business stakeholders, and two-weekly retrospectives are performed. As such the DUTCHCO market releases contain a combination of about 6 to 7 (bi-weekly) Sprint deliveries. As these Sprints do deliver working tested software, one could also call these releases. However, as these are only deployed in an acceptance test environment and not to

the market, we use the term 'market release' for those four releases each year.

The DUTCHCO teams are organized in a component-based way. One database-team (DB) is based in The Netherlands. Two teams are based in India; one portal and asset management-team of nine people (POR and AM), and one reporting-team (AR) of also nine people.

Table 1 summarizes the release approaches of both companies. BELTEL runs single projects that are combined eight times per year for user acceptance testing and deployment. In the DUTCHCO case no projects are to be found; user stories are combined in releases that are deployed every three months.

Table 1. Summary of Release Approaches

	BELTEL (N = 4)	**DUTCHCO (N = 1)**
Frequency of Release	6 weeks (8 releases per year)	3 months (4 releases per year)
Scope of Release	Collection of projects from different Business Domains (a mix of Scrum and plan-driven)	Collection of User Stories performed by 4 Scrum-teams (of which 3 offshore in India)
Average Size of Release	444 Function Points	776 Function Points
Average Cost of Release	2,190 K Euro	512 K Euro

3.3 Challenges in Comparing both Companies

Looking at the large differences in the project size, staff count, budgets, geographic location of team and customer demands we recognized major challenges in comparing software projects performed in a telecom company with a software company. To remedy this, we used a tool that we designed to address this challenge [29]. In previous research we built a model, the so-called *Cost Duration Matrix* (see Figure 2), based on the consideration that *Project Size*, *Project Cost*, *Project Duration* and the *Number of Defects* detected during a software project are interrelated with each other [5] [6] [7]. The model takes a project's size, measured in function points (FPs) [30], as starting point and as a source for normalization that makes it possible to compare software projects with different settings. The model compares the actual costs normalized to a function point (in Euros per FP) and duration (in days per FP) for a project of this size to benchmarked data, taken from a set of 492 finalized software projects in the financial and telecom application domains. This is done using two power regressions conducted on the 492 projects, permitting the computation of the 'expected' cost and duration of a project of a given size (measured in function points) [5] [7].

3.4 Metrics

In this paragraph we describe and explain the major metrics that are collected and analyzed for the subject projects.

3.4.1 Project Metrics

Four project metrics are collected on each project that is subject of the case study: *Project Cost* (in Euros), *Project Duration* (in months), and the *Number of Defects* found during the project. *Project Size* is measured in function points, according to the IFPUG industry standard [30]. Based on this, we determine the *Cost per Function Point*, *Days per Function Point*, and *Defects per Function Point*, using in each case the size in function points as weighting factor.

3.4.2 Estimation Quality Factor

The *Estimation Quality Factor* (EQF) is a measure of the deviation of a forecast to the actual cost or duration. EQF is a forecasting metric that depicts the quality of forecasts made during a project. The measure was defined by DeMarco [31]. He defines EQF by:

$$EQF = \frac{Area\ under\ actual\ value}{Area\ between\ forecast\ and\ actual\ value}$$

We use the formalization proposed by Eveleens and Verhoef [32]. We reiterate and correct the definition given there. Let a be the actual value ($a > 0$), t_a the time the actual is known and $e(t)$ the value of the forecast at time t ($0 \le t \le t_a$) in the project. Then, the EQF is represented by [32]:

$$EQF = \frac{\int_0^{t_a} a \, \mathrm{d}t}{\int_0^{t_a} |a - e(t)| \, \mathrm{d}t}$$

$$= \frac{\int_0^{t_a} 1 \, \mathrm{d}t}{\int_0^{t_a} |1 - e(t) / a| \mathrm{d}t}.$$

EQF allows us to quantify the quality of forecasts. A low EQF value means that the deviation of the forecasts to the actual cost or duration is large. EQF is measured for both cost and duration.

3.4.3 Cost Duration Index

The *Cost Duration Index* is a measure of the relative position of a project within the *Cost Duration Matrix* (see Figure 2). The index is represented as a number between zero and one hundred. In practice most projects score between 80 and 99. A high index corresponds to a good position in the *Cost Duration Matrix* (best is top-right in the *Good Practice* quadrant). The index is based on the geometric mean of two proportions comparing the actual value to the benchmark value:

$$p = \sqrt{\frac{Actual\ Duration}{Benchmark\ Duration} * \frac{Actual\ Cost}{Benchmark\ Cost}}$$

We subsequently normalize this p to a value ranging from 0-100 with 100 being best via:

$$Cost\ Duration\ Index = \frac{(p_{\max} - p)}{(p_{max})} * 100$$

3.4.4 Stakeholder Satisfaction

Human satisfaction is a complex concept, involving many components such as physical, emotional, mental, social, and cultural factors [33] [34]. From behavioral science and consumerism multiple theories have emerged on psychometrically validated surveys on satisfaction (e.g. [35] [36]). Although extended handbooks are available on the setup of a satisfaction survey [37], we opted for a lean survey setup. The main reason to do so was a requirement from BELTEL executives to make the survey as short as possible in order to minimize disturbance to the daily work for employees. An important argument for this requirement was the fact that the survey was implemented as an integral part within the release process, meaning that some staff members had to fill it out several times during the release process (e.g. release managers that where involved in more projects that were included in one release filled out a separate survey for each individual project), or within every release (e.g. team members of Scrum-teams). We assume that this light-weight requirement will apply for other companies too and therefore is a precondition for a successful metric.

Stakeholder Satisfaction is a measure of the satisfaction of stakeholders of a specific project with the way a project was performed and with the results as delivered by that project. *Stakeholder Satisfaction* is measured by asking stakeholders of a specific project to rate their satisfaction on two aspects; the way a project was performed (the project's process), and with the results as delivered by a project (the project's result), for which we use questions with a 1 to 5 rating scale.

In both BELTEL and DUTCHCO surveys were answered by internal stakeholders of projects; e.g. project managers, developers, testers, product owners. In case external stakeholders were included, these were working for BELTEL OR DUTCHCO as client or business analyst for a specific project. No external stakeholders in the meaning of end-users of a projects' deliverables were involved in the surveys.

3.4.5 Perceived Value

Value of software projects is a complex metric to measure [18], and studies are not specific on how they define value [16]. It is difficult, if not impossible, to measure objectively and indisputable the real value as delivered by software projects to customers of BELTEL and DUTCHCO. Is *real value* about money and time? Does it mean financial value, as in studies indicated by Return Of Investment (ROI) [38]? Or is *real value* measured by Net Promotor Score (NPS), as other studies indicate [39] [40] [41]? Such holistic measurements on value are often difficult to make for a single project, and they cannot easily be related to single software projects, mainly because too many different factors are of influence for such measurements.

To approximate the real value, we measure *Perceived Value* as a qualitative measure of the perception of stakeholders of each project. This is based on the notion that in fact every measurement is an agreement on a measurement procedure that sufficiently approaches the actual value [38].

Perceived Value is measured for each stakeholder in a specific project, on four aspects: BELTEL's or DUTCHCO's customers, BELTEL or DUTCHCO's financials, BELTEL or DUTCHCO's internal process effectiveness, and BELTEL or DUTCHCO's innovation. We base the use of the four perspectives *Customer*, *Financial*, *Internal Process*, and *Innovation* on the Balanced Scorecard [42]. Based on the results per project of the four perceived value measures a *Perceived Value (overall)* is calculated, with the number of measures (not counting the choice "Don't know") as weighting factor.

3.5 Project Selection

Because we are particularly interested in data of finalized projects, all metrics are measured once a release is finalized, since only then we know the actual cost and duration of projects. Since we want to measure the effects of *Stakeholder Satisfaction* and *Perceived Value* on a software portfolio as a whole, we did not make any selection in the subset of projects within each release, except for the fact that we only selected projects that delivered software functionality (the projects could be counted in function points). Projects that do not include any software component (e.g. infrastructure projects or configuration projects) are excluded from our study.

3.6 Data Collection procedure

3.6.1 Collection of quantitative data

Within BELTEL, a major part of the data collection for our case study was performed within the measurement capability that was already operational within the software department of the organization. Data collection on *Project Cost*, *Project Duration*, *Number of Defects*, *Project Size*, and calculation of both *Estimation Quality Factor* metrics was performed by members of a measurement team that was supported (for performing function point counts [30]) by measurement staff of BELTEL's main Indian supplier.

Different artifacts were used as a source for function point counting, depending from the availability per project (e.g. sets of functional documentation, user stories recorded in one of the Scrum backlog tools, architectural documents, project documentation, user manuals, or wireframes). All project data was stored in a measurement repository that was provided for our study. The lead author of the study was part of the BELTEL's measurement team.

In the DUTCHCO case, a dedicated research project was performed in order to collect and analyze data of software releases. The lead author of this paper performed the size calculations in retrospect for DUTCHCO. Due to this, it was possible to replicate the study that we performed within BELTEL in exactly the same way in the DUTCHCO organization. All quantitative data was defined and collected in the same way. Function points were counted according to the same counting rules as used within BELTEL [30]. As a source for function point counting the user stories as recorded in the Scrum backlog tool were used.

Driven by the observations in our original study on correlations between *Project Cost* and *Number of Defects* on one hand, and *Stakeholder Satisfaction* on the other, we decided to collect data from finalized software releases within DUTCHCO in a more detailed way: cost data was categorized into a limited number of cost categories (e.g. design, build, test,

deploy, management overall, quality assurance), and defect data was collected per defect severity (e.g. blocking, critical, high, medium, low).

3.6.2 Collection of qualitative (survey) data

Besides the project data that was collected as an operational practice, we collected data on *Stakeholder Satisfaction* and *Perceived Value*. To do so we conducted a questionnaire with stakeholders from BELTEL, and later from DUTCHCO. The list of stakeholders was prepared in cooperation with the project managers of the applicable software projects, and consists of a mix of business and IT representatives that were involved in the subject projects. We asked the participants, who are stakeholders of a specific software project within a release, to rate their satisfaction with the way the project was performed and to rate their perception of the value that was added by the project. Besides ratings on a 1-5 rating scale we asked the participants to add free format text as an explanation of their perceptions. The questionnaire consists of five questions:

1. What was your role in project PROJECT_NAME?
2. How satisfied are you with the way project PROJECT_NAME was performed (the project's process)? (1-5 rating scale);
3. How satisfied are you with the results of project PROJECT_NAME (the results as delivered by the project)? (1-5 rating scale);
4. How would you rate the delivered value of project PROJECT_NAME to the following aspects (1-5 rating scale, with 'Don't know' as an option; this choice was excluded from further analysis)?
   a. BELTEL's Customers (Value in terms of delivered to customers of BELTEL);
   b. BELTEL Financial (Value in terms of financial revenue for BELTEL);
   c. BELTEL Internal Processes (Value in terms of improvement and/or proper performance of BELTEL's internal processes);
   d. BELTEL Innovation (Value in terms of innovation of BELTEL's products or services delivered to its customers)?
5. Are there any additional comments or suggestions you'd like us to know about this project? (Free format text).

With regard to question 4: the additional information (between brackets) was shown to the participants when hovering with a mouse pointer over a question mark next to the text of each of the four aspects.

Within DUTCHCO we applied the same electronic survey for stakeholders of the finalized software releases, including team members from the INVEND teams located in India.

3.7 Analysis Procedure

To explore potential relationships between the collected metrics, we tested for association between paired samples. Because all sample data is not normally distributed (see Table 3 for details on skewness and kurtosis and the boxplots in Figure 1), we used a Spearman rank correlation coefficient test for this purpose. In order to understand the underlying principles that can explain the outcomes of the quantitative analysis, we studied the free format text from the surveys.

Following Hopkins [43], we prevent from Type I errors, e.g. finding a correlation by chance, simply because multiple comparisons are performed on the same dataset, by performing Benjamini-Hochberg corrections on all p-values. We used an alpha of 0.05/26 (the number of projects in scope of this study), meaning that we assume all p-values above 0.0019 as not significant [43]. We consider a significant correlation higher than 0.3 (or lower than −0.3) to be moderate, a significant correlation score higher than 0.5 (or lower than −0.5) to be strong, and a significant correlation above 0.9 (or lower than -0.9) to be very strong.

To compare the outcomes of the quantitative analysis of the project metrics with the survey we coded the free format text that resulted from the surveys that were performed within BELTEL and DUTCHCO. We use the tool Qualyzer[2] for this purpose. We applied open coding, breaking down the survey data into first level concepts and second-level categories. Coding was performed by the first author of the study, and reviewed by the other authors.

4. RESULTS

4.1 Description of the BELTEL Projects

Within the scope of our study we evaluated four software releases within BELTEL, covering a total of 22 software projects. Table 2 gives a brief description of each project, where the numbering of the projects indicates in which release each project was finalized and in which company a project or release was performed (e.g. BELTEL 6.4 is a BELTEL project that finalized in Release 6).

The software projects in scope represent a varied outline of BELTEL's software project portfolio. It includes projects of different business domains, sizes, cost patterns, durations, and delivery approaches. Some projects are typically once-only, with teams that were put together for the purpose of one project only. Others are part of subsequent iterations within a release structure with a steady heartbeat and a fixed, experienced team. Sixteen projects are characterized as plan-driven, while six followed a more agile (Scrum) delivery approach, however a formal Scrum-by-the-book approach was not in place (i.e. sprints where performed, a backlog was managed and prioritized in a backlog tool, a product owner was in place, however no retrospectives were performed, no Scrum master was in place).

All projects were performed separately. Yet from the User Acceptance Testing onwards they were combined as a release deployed into BELTEL's production environment. Looking at the total cost of a release, on average 60% was spent on software projects. The remaining cost were spent on infrastructure projects, small innovations, and configuration projects, and as such do not fit into the *Cost Duration Matrix* approach. These projects are out-of-scope for this case study.

[2] http://qualyzer.bitbucket.org

Table 2. The BELTEL projects in scope of the case study.

Project ID	Project Description
BelTel 3.1	Rules- and regulations driven small Billing project
BelTel 3.2	Implementation of a control on a Billing application
BelTel 3.3	Release-based enhancements on CRM-application (Scrum)
BelTel 3.4	New campaign management tool (3rd part of a program)
BelTel 3.5	Release-based enhancements on a mobile App (Scrum)
BelTel 4.1	Enhancements on a Billing application
BelTel 4.2	Release-based enhancements on CRM-application (Scrum)
BelTel 4.3	Front end project: Connect Google Play
BelTel 4.4	Rules & Regulations enhancement: fee for customers
BelTel 5.1	Release-based enhancements on CRM-application (Scrum)
BelTel 5.2	New campaign management tool (4th part of a program)
BelTel 5.3	Data warehouse 4 sprints of an iteration (Scrum)
BelTel 6.1	Enhancement to integrate payment by credit-card-aliases
BelTel 6.2	Enhancement to implement Apple Store code
BelTel 6.3	Release-based enhancements on CRM-application (Scrum)
BelTel 6.4	Adapt a procedure on an online platform
BelTel 6.5	E-invoice for a subset of customers in a Billing system
BelTel 6.6	Easy Script for cleanup of master MSISDN
BelTel 6.7	Rules & Regulations project on a Billing application
BelTel 6.8	Front end enhancement: Shopper user interface e-services
BelTel 6.9	Once-only migration project
BelTel 6.10	New Order Management System (part of program, Scrum)

All project data of BelTel is to be found in a Technical Report [73].

Table 3 gives an overview of the descriptive statistics of the BELTEL projects involved in the case study. As the table shows, the software projects in scope of the BELTEL study are all relatively small in size, when compared to the projects in our research repository, ranging from 4 to 4600 Function Points (FPs): *Project Size* ranges from 12 FPs to 324 FPs, with a median of 39 FPs.

To examine differences between the BELTEL projects in scope of this study with our research repository as a whole, holding data of 492 software projects from four different companies, we performed Wilcoxon ranked sum tests with Bonferroni corrections to compare overall differences, and differences per size (see Table 4).

Table 3. Descriptive statistics of the BELTEL project data.

	Project Data (n = 22)				
	Cost Duration Index	Project Cost (EUR)	Project Duration (Months)	Project Size (FPs)	Number of Defects
Minimum	86.92	8,000	4.96	12	1
First Quartile	93.27	44,001	8.37	25	3
Median	97.28	66,209	10.18	39	9
Third Quartile	98.57	118,876	11.73	126	23
Maximum	99.78	296,000	19.03	324	223
Mean	95.90	99,615	10.20	79	29
Skewness	-1.03	1.27	0.78	1.71	3.19
Kurtosis	0.06	0.77	1.43	2.71	10.89
St. Deviation	3.51	78,209	3.22	82	55

Table 4. Results from a Wilcoxon rank sum comparison of BELTEL releases (n = 22) with peer groups (n = 492).

	Median BelTel	Median peer group	W	p-value
Project Size	39	116	7148	0.0108
Project Cost	66,209	278,156	8003	0.0001
Project Duration	10.18	8.41	4008	0.0394
Number of Defects	9	72	2989	0.1028
Cost per FP	1612	2520	6952	0.0239
Days per FP	7.85	2.08	3058	0.0006
Defects per FP	0.22	0.16	2280	0.6621

The in light grey highlighted rows indicate statistically significant difference when applying Bonferroni corrections based on 22 comparisons, at the overall level of significance of 0.05 (we assumed all p-values above 0.0023 as not significant).

If the data were sampled from a population with the median of the research repository, one would expect the sum of signed ranks (in the table reported as W) to be relatively small. The comparison shows that BELTEL significantly differs from the other projects in the repository on *Project Cost*, as well on *Days per FP*. On all other metrics no significance was found in the test. With regard to Project Cost we see this effect also in the boxplot in Figure 1; BELTEL clearly shows overall lower cost for its projects compared to the other companies in our research repository. Although not confirmed by the statistical tests, a similar effect can be seen for Project Size; the boxplots in Figure 1 indicate that overall size for BELTEL projects is smaller than that of the other companies. An explanation for differences in the outcomes of statistical tests and the boxplots in Figure 1 might be that the first only includes the 22 BELTEL projects that are in scope of this study, while the second includes all 157 BELTEL projects from our research repository.

Besides project metrics, we collected data of the BELTEL projects on *Stakeholder Satisfaction* and *Perceived Value* by sending an online questionnaire to applicable stakeholders of each software project once the technical go live was performed. The overall completion rate of all surveys within BELTEL was 69%. Over a period of four releases 103 surveys were completed by 53 individual respondents. One respondent could answer surveys for different projects in one release, or repeated surveys for a series of iterative projects over different releases.

4.2 Description of the DUTCHCO projects

Within the scope of our study we examined four DUTCHCO releases, all built from a large number of user stories. Table 5 gives a brief description of each release, where the numbering of the releases indicates in which company a release was performed; e.g. DUTCHCO 5.1 AM is a DUTCHCO release that was applicable to the asset management (AM) component of its billing solution.

Table 5. The DUTCHCO projects in scope of the case study.

Project ID	Project Description
DutchCo 5.1 AM	Release containing asset management (AM) user stories.
DutchCo 5.1 POR	Release containing portal (POR) user stories.
DutchCo 5.2	Release applicable on DutchCo's general billing solution.
DutchCo 5.3	Release applicable on user stories for customer VF/BK.

All project data of DutchCo is to be found in a Technical Report [73].

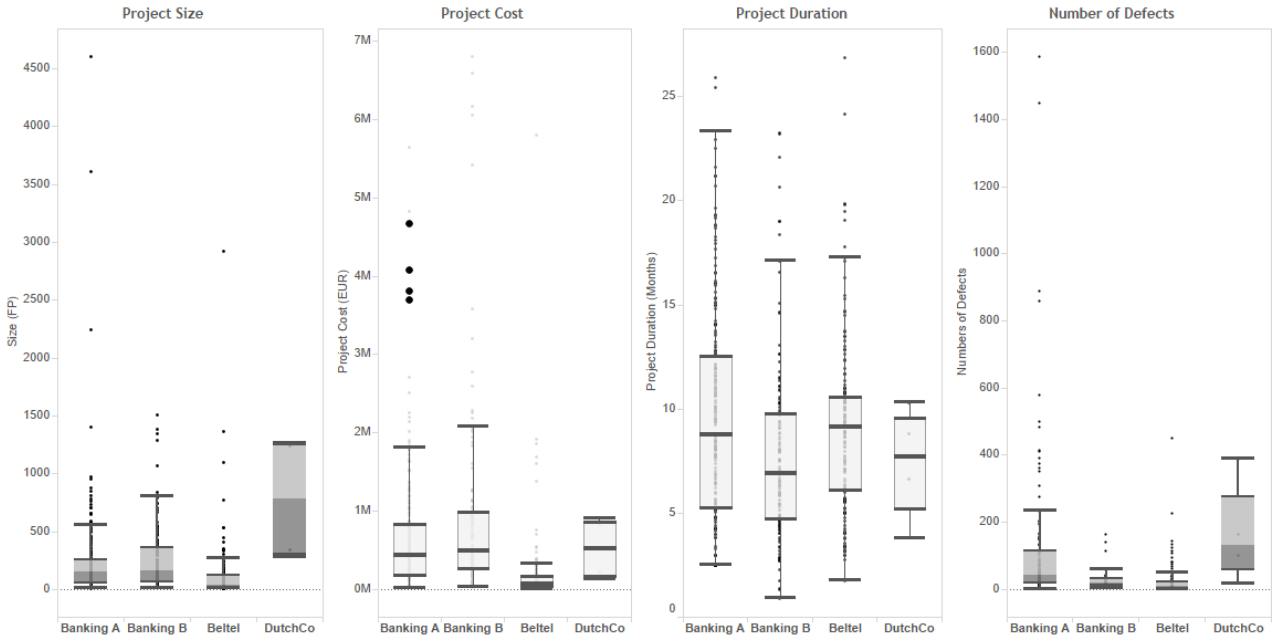


Figure 1. Boxplots of resp. *Project Size, Project Cost, Project Duration,* and *Number of Defects* of four organizations that are incorporated in our research repository of 492 projects. The boxplots indicate that DUTCHCO projects significantly deviate on *Project Size* and *Number of Defects* from projects from other companies in our research repository, and not as such on *Project Cost* and *Project Duration*.

Unlike BELTEL, where the software portfolio includes a mix of projects of various business domains, delivery models and governance structures, the portfolio of DUTCHCO is more heterogeneous in nature. DUTCHCO implements only four releases each year to its customers. Due to that, these releases are usually quite large in size. All DUTCHCO releases relate to the same business domain, namely the billing solution it provides to its customers. However, it occurs that different sets of functionality are delivered to customers, due to differences in requirements.

DUTCHCO's user stories are maintained in its backlog management tool, and continuously bundled in sprint backlogs. As a result, the governance structure of DUTCHCO is relatively simple. There are no projects, and there is a limited budget and planning activity. DUTCHCO has adopted a Scrum approach. Scrum teams are organized by functional component (e.g. Portal, Asset Management, Reporting, and Database). A large part of the Scrum teams is working from India, managed by INVEND.

Table 6 gives an overview of the descriptive statistics of the four DUTCHCO releases involved in this extended case study. The DUTCHCO subset exists of four releases, two relatively smaller ones (although still as large as BELTEL's largest projects), and two large ones. As can be seen in Figure 2 and in Table 6 above, the two oldest DUTCHCO projects (DutchCo 5.1 - POR and DutchCo 5.1 – AM), are smallest in size (resp. 277 and 335 FPs). However, both are comparable to the largest projects from the BELTEL case. The two newest DUTCHCO releases are relatively large, compared to the projects in the

Table 6. Descriptive statistics of the DUTCHCO project data.

	Project Data (n = 4)				
	Cost Duration Index	Project Cost (EUR)	Project Duration (Months)	Project Size (FPs)	Number of Defects
Minimum	95.03	125,827	3.78	277	15
First Quartile	96.12	192,040	5.90	321	78
Median	96.81	514,486	7.71	784	131
Third Quartile	97.41	835,343	9.18	1240	219
Maximum	98.22	896,788	10.29	1261	386
Mean	96.72	512,987	7.37	777	166
Skewness	-0.147	-0.002	-0.209	-0.003	0.434
Kurtosis	-1.952	-2.404	-2.054	-2.432	-1.865
St. Deviation	1.33	399,020	2.832	543	158

Table 7. Results from a Wilcoxon rank sum comparison of DUTCHCO releases (n = 4) with peer groups (n = 492).

	Median DutchCo	Median peer group	W	p-value
Project Size	784	116	219	0.0074
Project Cost	514,486	278,156	808	0.5387
Project Duration	7.71	8.41	1135	0.5981
Number of Defects	131	72	240	0.0564
Cost per FP	686	2520	1791	0.0047
Days per FP	0.28	2.08	1789	0.0048
Defects per FP	0.22	0.16	565	0.8762

The in light grey highlighted rows indicate statistically significant difference when applying Bonferroni corrections based on 4 comparisons, at the overall level of significance of 0.05 (we assumed all p-values above 0.0125 as not significant).

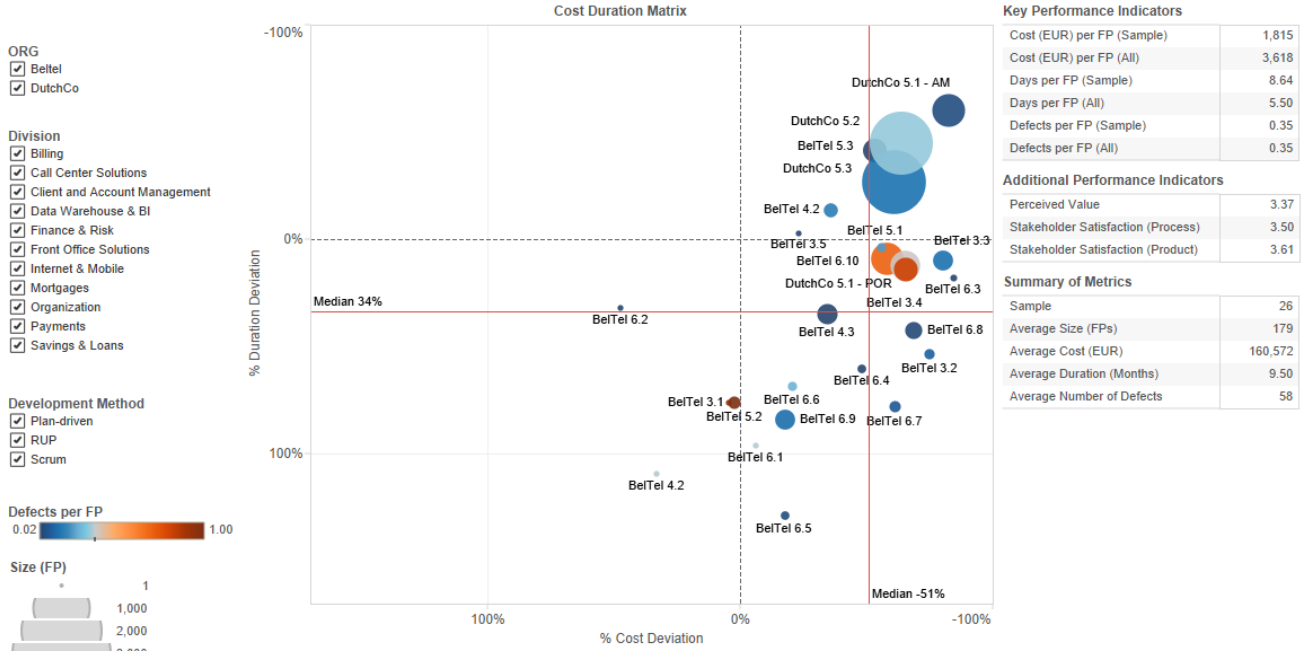


Figure 2. A Cost Duration Matrix showing the 22 BELTEL and 4 DUTCHCO projects that are subject of the study.

BELTEL subset: DutchCo 5.2 is 1233 FPs in size, and DutchCo 5.3 is 1261 FPs in size. Apparently driven by a schedule of four deployments per year, combined with a tendency to bundle the user stories of all functional components of its system, DUTCHCO releases tend to grow relatively large.

We performed a Wilcoxon rank sum comparison with Bonferroni corrections between the DUTCHCO subset and our research repository as a whole, holding data of 492 software projects from four different companies, to compare overall differences, and differences per size (see Table 7 and Figure 1). The comparison shows that DUTCHCO significantly differs from the other projects in the repository on *Project Size*, as well on *Cost per FP* and *Days per FP*. On all other metrics no significance was found in the test.

We observe two findings here. Firstly, DUTCHCO releases have on average a larger size than other projects in our repository, which is good. This leads to a positive effect, from benchmarking purposes; due to the larger size of DUTCHCO releases also *Cost per FP* and *Days per FP* are better than the values of the other companies in our research repository.

Although no statistical evidence is found for any differences between the *Number of Defects* of both distributions (see Table 7), the boxplot view in Figure 1 indicates that besides *Project Size*, DUTCHCO also deviates from its peer groups on *Number of Defects*. Based on this we assume that a good score on *Project Size* might be counterbalanced here by a bad score on *Number of Defects*.

Besides project metrics as described above, we collected data on *Stakeholder Satisfaction* and *Perceived Value* by sending an online survey to applicable stakeholders of each software release once the technical go live was performed. For this purpose, we used the same electronic survey that was used before within BELTEL. The overall completion rate of all surveys within DUTCHCO was 71%. Thirty (30) surveys were completed by 30 individual respondents of both DUTCHCO and INVEND. Due to the fact that the three first releases were measured relatively long after finalization of each release, only for the latest DUTCHCO release an electronic survey was performed.

4.3 *Results of plotting on the Cost Duration Matrix*

We used the model that we developed in previous research to compare a portfolio of projects to the benchmark, by means of a *Cost Duration Matrix* [5] [7], as shown in Figure 2 for the 26 projects under study in this paper. Each project is shown as a circle. The larger the circle, the larger the project is (in function points), and the 'redder' the project is, the more defects per function point it contains. The position of each project in the matrix represents the cost and duration deviation of the project relative to the benchmark, expressed as percentages. The horizontal and vertical 0%-lines represent zero deviation, i.e. projects that are exactly consistent with the benchmark. A project at (0%, 0%) would be one that behaves exactly in accordance with the benchmark; a project at (-100%, -100%) would cost nothing and be ready immediately; and a project at (+100%, +100%) would be twice as expensive and take twice as long as expected from the benchmark.

As can be seen from the figure, most of the 26 projects in the portfolio are cheaper than the benchmark would predict (right of the 0%-cost bar), yet take longer than expected (below the 0%-duration bar). The 0%-lines divide the *Cost Duration Matrix* into four quadrants:

Table 8. Matrix with test results of association between paired samples, using Spearmans's rank correlation coefficient.

	Project Size	Project Cost	Project Duration	Number of Defects	Cost per FP	Days per FP	Defects per FP	Cost Duration Index	Stakeholder Satisfaction (Process)	Stakeholder Satisfaction (Result)	Perceived Value (Overall)	Perceived Value (Customer)	Perceived Value (Process)	Perceived Value (Financial)	Perceived Value (Innovation)	Estimation Quality Factor (Cost)
Project Cost	0.81 (0.000)															
Project Duration	-0.13 (0.53)	0.09 (0.65)														
Number of Defects	0.68 (0.000)	0.70 (0.000)	-0.22 (0.32)													
Cost per FP	-0.70 (0.000)	-0.19 (0.35)	0.35 (0.08)	-0.36 (0.10)												
Days per FP	-0.96 (0.01)	-0.72 (0.000)	0.35 (0.08)	-0.66 (0.001)	0.73 (0.000)											
Defects per FP	-0.14 (0.54)	0.10 (0.67)	0.05 (0.83)	0.59 (0.004)	0.39 (0.07)	0.15 (0.50)										
Cost Duration Index	0.09 (0.66)	-0.03 (0.88)	-0.01 (0.95)	0.18 (0.43)	-0.24 (0.22)	-0.06 (0.76)	0.03 (0.89)									
Stakeholder Satisfaction (Process)	0.02 (0.93)	-0.26 (0.24)	-0.44 (0.04)	-0.30 (0.23)	-0.36 (0.10)	-0.21 (0.34)	-0.50 (0.04)	0.01 (0.97)								
Stakeholder Satisfaction (Result)	-0.00 (0.99)	-0.16 (0.48)	-0.47 (0.03)	-0.07 (0.78)	-0.27 (0.22)	-0.21 (0.35)	-0.23 (0.35)	-0.04 (0.86)	0.72 (0.000)							
Perceived Value (Overall)	0.32 (0.15)	0.07 (0.75)	-0.10 (0.68)	0.04 (0.89)	-0.34 (0.13)	-0.37 (0.10)	-0.20 (0.44)	0.04 (0.86)	0.09 (0.71)	0.24 (0.30)						
Perceived Value (Customer)	0.34 (0.13)	0.10 (0.68)	-0.09 (0.69)	0.06 (0.82)	-0.35 (0.12)	-0.38 (0.09)	-0.20 (0.43)	0.04 (0.85)	0.07 (0.77)	0.23 (0.31)	0.99 (0.000)					
Perceived Value (Process)	0.26 (0.26)	0.01 (0.98)	-0.03 (0.89)	-0.06 (0.82)	-0.27 (0.24)	-0.29 (0.20)	-0.15 (0.56)	0.06 (0.78)	0.10 (0.66)	0.20 (0.37)	0.98 (0.000)	0.96 (0.000)				
Perceived Value (Financial)	0.32 (0.15)	0.07 (0.75)	-0.10 (0.68)	0.04 (0.89)	-0.34 (0.13)	-0.37 (0.10)	-0.20 (0.44)	0.04 (0.86)	0.09 (0.71)	0.24 (0.30)	1.00 (0.000)	1.00 (0.000)	0.98 (0.000)			
Perceived Value (Innovation)	0.34 (0.13)	0.10 (0.68)	-0.09 (0.69)	0.06 (0.82)	-0.35 (0.12)	-0.38 (0.09)	-0.20 (0.43)	0.04 (0.85)	0.07 (0.77)	0.23 (0.31)	0.99 (0.000)	1.00 (0.000)	0.94 (0.000)	0.99 (0.000)		
Estimation Quality Factor (Cost)	-0.08 (0.78)	0.02 (0.98)	-0.03 (0.92)	-0.32 (0.37)	0.11 (0.74)	-0.03 (0.93)	-0.28 (0.43)	0.13 (0.68)	-0.02 (0.96)	0.16 (0.62)	-0.23 (0.51)	-0.23 (0.51)	-0.23 (0.50)	-0.23 (0.51)	-0.23 (0.51)	
Estimation Quality Factor (Duration)	-0.36 (0.10)	-0.43 (0.05)	-0.30 (0.01)	-0.18 (0.45)	0.16 (0.48)	-0.23 (0.31)	0.00 (0.99)	-0.25 (0.26)	0.35 (0.12)	0.20 (0.38)	-0.32 (0.16)	-0.35 (0.13)	-0.28 (0.24)	-0.32 (0.16)	-0.36 (0.13)	-0.35 (0.26)

The table above shows results from a test of association between paired samples of the 26 software projects from both case studies, using Spearman's rank correlation coefficient. Due to the fact that the DUTCHCO case contained a limited number of four projects, we performed the association tests over the total set of 26 projects from both BELTEL and DUTCHCO. The overview shows for each test the correlation coefficient and between brackets the p-value. We counteracted the problem of multiple comparisons by performing a Benjamini-Hochberg correction [69] [70] [71]. We compared each individual p-value to its Benjamini-Hochberg critical value, (i/m)Q, where i is the rank, m is the total number of tests, and Q is the false discovery rate. We used 0.10 as false discovery rate, according to [68]. The largest p-value that has $p<(i/m)Q$ is significant, and all of the p-values smaller than it are also significant, even the ones that aren't less than their Benjamini-Hochberg critical value. A color indicates samples that are correlated: dark grey indicates a very strong (positive or negative) linear relationship (correlation coefficient higher than 0.70), moderate grey indicates a strong linear relationship (correlation coefficient between 0.50 and 0.70). Significant samples with a correlation coefficient lower than 0.50 are indicated in light grey.

1. *Good Practice* (top right); projects that score better than average for both cost and duration. In Figure 2, there are six projects in this quadrant, of which three of BELTEL (5.3, 4.2, and 3.5) and three of DUTCHCO (5.1 - AM, 5.2, and 5.3).
2. *Cost over Time* (bottom right); projects that score better than average for cost, yet worse than average for duration. This is where the majority of projects are in Figure 2.
3. *Bad Practice* (bottom left); projects that score worse than average for both cost and duration. In Figure 2, there are four projects in this quadrant, all from BELTEL.
4. *Time over Cost* (top left); projects that score better than average for duration, yet worse than average for cost. In Figure 2, there are no projects in this quadrant.

The overall performance of the portfolio is furthermore summarized through the two red 'median' lines: On average, projects in the subject portfolio take 34% more time than expected from the benchmark, yet are 51% cheaper. The *Cost Duration Matrix* provides a tool to compare two project portfolios in terms of *Project Cost* and *Project Duration*. Our comparisons are based on the benchmark of 492 projects from the finance and telecom industries, described in more detail in [5] [7]. The benchmark of 492 projects contain 157 previous

projects from BELTEL, and 4 previous ones from DUTCHCO, making it a suitable benchmark to compare the new additional 26 projects against.

4.4 Results of the tests for association

To identify potential relationships between the different metrics that we collected we performed a series of tests on paired samples of each metric, by using Spearman rank correlation coefficient. Because for only one DUTCHCO release data on *Stakeholder Satisfaction* and *Perceived Value* was measured (for only the latest release a survey was performed), we decided to test for associations on the BELTEL and DUTCHCO dataset as a whole. The results of these tests are shown in Table 8. The table is setup in the form of a matrix that pairs sets of two metrics. For each pair the correlation coefficient is shown, including (between brackets) the associated p-value. A color indicates correlation: dark grey indicates a strong (positive or negative) linear relationship, bright grey indicates a moderate linear relationship, light grey indicates a weak linear relationship. Results of the tests for association on the BELTEL projects only can be found in Table 3 of the original research paper [6].

However, a remark on the way we interpreted the results in Table 8 is in place. If results relating to the previous BELTEL analysis agree with the results including the DUTCHCO data, we assume that both organizations are exhibiting similar results. If the results are completely different when the DUTCHCO results are included, we conclude that the companies are behaving differently, and that further research is needed to establish whether the new combined results are valid.

A second warning is in place with regard to some of the metrics we use. As it is dubious practice to correlate metrics that have a functional relationship between them (e.g. *Cost per FP* and *Days per FP*), as likely spurious correlations are found [44], we do not valid any findings with regard to these metrics as reliable.

Analysis of the statistical tests for association between paired samples as depicted in Table 8 results in the following observations.

Observation 1: Strong positive correlations are found between Project Size, Project Cost, and Number of Defects.

In the first column of Table 8, it can be seen that *Project Size*, measured in function points, is strongly associated with *Project Cost* and *Number of Defects*. This effect is known from related studies [45] [5] and as such not a surprise in our research. The second column shows that also among themselves *Project Cost* and *Number of Defects* are strongly interrelated. However, where in many other organizations a clear correlation is found between *Project Size* and *Project Duration*, both BELTEL and DUTCHCO show an atypical pattern. *Project Size* and *Project Duration* are not related in any way.

This is the case when both BELTEL and DUTCHCO are analyzed in a combined way, like inventoried in Table 8 and plotted in Figure 3, and also when examined separately (though for DutchCo's four data points we have no statistical significance). In order to examine whether this effect is only linked to the set of 26 projects in scope, or whether this effect goes for BELTEL as a whole, we perform the test also with the BELTEL projects that are not included in this paper, yet available in our repository. A test with all 157 BELTEL projects included shows a p-value of 0.002, and a correlation coefficient of 0.24, indicating that also in this case no correlation between Size and Duration is found. This outcome supports our observation that regardless the size of a project the duration is typically ten months. This is confirmed by a relatively low standard deviation for BELTEL's *Project Duration* (see Table 3).

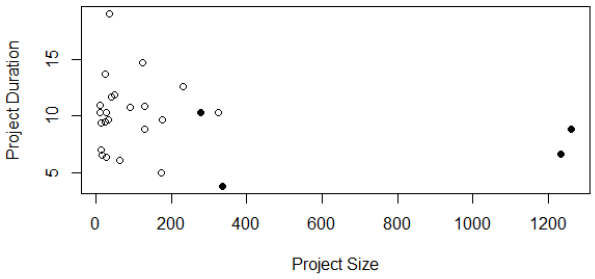


Figure 3. Plot of Project Duration versus Project Size; BELTEL projects are indicated in open dots, DUTCHCO in closed dots.

In spite of this atypical effect with regard to *Project Duration*, in the fifth row a strong correlation can be seen between *Days per FP* and *Cost per FP*. Besides that we observe a relation between *Days per FP* and *Project Size* and *Project Cost*. However, due to the functional relationship between both metrics we do not valid these findings as reliable [44].

Observation 2: Stakeholder Satisfaction for both process and result are strongly interrelated to each other. Stakeholder Satisfaction relates negatively with Project Duration.

Row nine of Table 8 shows that an observation that was already found in our original study [6] with regard to *Stakeholder Satisfaction*, remains intact. Both satisfaction ratings for process and product correlate strongly with each other. The fact that the same results are found for BELTEL alone, and also when the DUTCHCO data is added provides evidence that the observation applies to both companies and may represent a more general observation that high satisfaction ratings on process link with high ratings on the delivered product. However, the weak correlation between Project Duration and Costs per FP and Days per FP was not visible in the BELTEL data and has only occurred with the addition of the DUTCHCO data. This suggest the effect is due to the DUTCHCO data (based on a single large release) and calling for more research to investigate whether the effect is real.

Column three shows that *Project Duration* has a moderate negative relation with *Stakeholder Satisfaction* for both Process and Result. Longer project durations tend to lead to lower satisfaction rates. Furthermore Project Duration relates weakly with Cost per FP and Days per FP, indicating that longer project

durations lead to higher Cost per FP and a higher number of Days per FP. However, due to the functional relationship between both metrics we do not valid these findings as reliable [44].

Two observations are related to *Perceived Value*. A weak uphill linear relationship between *Project Size* and *Perceived Value (overall)*, as shown in our original study, is found here too, indicating that *perceived value* is higher for larger projects (in Function Points). Furthermore, several of the perceived value metrics shows weak negative relations with *Cost per FP* and *Days per FP*, indicating that lower cost and duration per FP links with higher scores on perceived value. This effect is much reduced compared with the original study with BELTEL data only [6], suggesting that it is a BELTEL phenomenon.

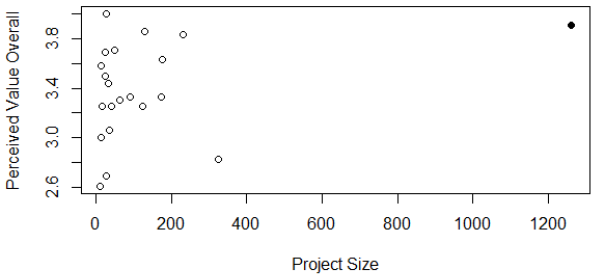


Figure 4. Plot of Perceived Value Overall versus Project Size; BELTEL projects are indicated in open dots, DUTCHCO in closed dots.

A major limitation here is, that the DUTCHCO project for which *Perceived Value* and *Stakeholder Satisfaction* is measured, is significantly larger in size than all other BELTEL projects. Figure 4, with on the X-axis the Project Size in Function Points, and on the Y-axis the Overall Perceived Value rating of each project, clearly shows that a good comparison in fact is not yet possible in this context; more data is needed, especially from relatively larger projects. An additional remark on this phenomenon is that it may be that the correlations would have been weaker for BELTEL in the original paper if that analysis had been based on a more robust correlation coefficient.

We observe a striking correlation between all mutually *Perceived Value* measurements. We assume that the four aspects are measuring the same construct, or that the answers to those items were influenced in the same way. This effect was not measured this strongly with BELTEL data only. We assume the effect found now is an artefact of adding DUTCHCO data. Due to the fact that the results are unstable, we do not value these outcomes to high though.

Other observations with regard to perceived value, as mentioned in our original paper [6], seem to have vanished in this study. After adding the DUTCHCO data to the comparison, no relations with another project metric are observed.

A comparison of the results of the test for association which listed only BELTEL results (see Table 3 in the original study [6]), with the results of the test in which both BELTEL as DUTCHCO projects are included (see Table 8), shows that the latter shows a clearer and more coherent pattern. Where the original, BelTel only, table shows a rather scattered pattern, the actual results focus on the three observations mentioned above. Especially the statistical power of function points as a measure for *Project Size* stands out. Besides that, we found indications for a positive relationship between both *Stakeholder Satisfaction* measures, and between *Stakeholder Satisfaction* for results and *Project Duration*. We did not find direct evidence for strong relations between *Perceived Value*. However, we do have expectations with regard to this for future research due to a very strong interconnection between the four *Perceived Value* measures. In the next paragraph we challenge our observations by linking them to the free format text that resulted from the surveys that are performed at closure of each release.

Observation 3: Weak correlations are found between Estimation Quality Factor for Duration on the one hand and Project Duration and Stakeholder Satisfaction on the other.

A final observation that results from the quantitative analysis is about the quality of estimations with regard to project duration (see the bottom horizontal row in Table 8). When compared to the initial BELTEL study, the only consistent observations are the negative correlation between EQF (Duration) and Project Duration and the positive correlation between Stakeholder Satisfaction and EQF (Duration). The first correlation suggests that shorter projects are less well estimated with regard to duration. However, this effect was not visible in the analysis of the BELTEL data, so must be due to the DUTCHCO data. The second suggests that stakeholders like accurate duration estimates, although in the initial study stakeholders were satisfied about the result, while after adding data stakeholders were satisfied about the process.

4.5 Results of the free format text analysis

In order to compare the outcomes of the quantitative analysis of the project metrics with the survey we coded the free format text that resulted from the surveys that were performed within BELTEL and DUTCHCO. See Table 9 and Table 10 for the outcomes of the coding of BELTEL and DUTCHCO free format text data. Both tables are ordered on the number of times a code was applied in the comments. We discovered seven main themes: In the following paragraphs we discuss these main themes, where we combined connected coding aspects into one theme. A subset of comments given by participants from the surveys is included in the following paragraphs, indicated by the letter "B" (for BELTEL) or "D" (for DUTCHCO) followed by a participant number.

4.5.1 Quality, Deployment and Testing (A1, A3, A7)

The first thing that strikes us when looking at the results of the coding process is that aspects with regard to quality are high on the list of items that apply to the stakeholders. Most remarks were about good quality, however, a number had to do with low quality issues of deliverables.

A large number of negative comments given in the survey was related to the deployment of projects within a release into BELTEL's or DUTCHCO's production environment. Most had to do with issues that occurred during this process (e.g. problems with environments or incidents in production that needed to be fixed, repeated rollback of releases, improvements to be made in the deployment process, and in solving issues.

An explanation for the fact that many issues occur after going technically live is that BELTEL uses the first week (or sometimes a longer period) to test deployments in the production environment. Usually projects are not commercially live during that period. Comments with regard to testing are often related to these deployment issues. Also here we find a majority of comments that are related to issues with test environments and the test process itself, for example:

'A lot of discussion on how we need to test...' (B39).

'There were some defects in production' (D30).

We note that deployment itself is not mentioned by any of DUTCHCO's stakeholders. Maybe an explanation for this is the fact that (unlike the BELTEL approach with centralized deployment by a separate release team) DUTCHCO teams are themselves responsible for deployment of solutions. Summarizing, for both BELTEL and DUTCHCO a generic observation applies on quality:

Observation 4: Satisfied stakeholders tend to emphasize good quality, while dissatisfied stakeholders say testing and deployment need improvements.

4.5.2 Communication (A2)

The second most mentioned point on the stakeholder's list in both studied companies is about communication. A number of remarks have to do with good communication between parties. A remarkable finding within BELTEL was that positive remarks all were related to external suppliers in the frontend development of website and app development, and not with the main strategic supplier INDSUP.

In the DUTCHCO case many positive remarks on commination had to do with team aspects, such as:

'Improvements in the cooperation between teams. A lot of work done in a short amount of time' (D02).

However, not all is well with communication. Besides the many positive remarks, there are also suggestions for improvement, sometimes related to the agile process:

'Communication and involvement for agile items is limited to the bare minimum, so the added value of release management is not really large here. The whole agile process is still pretty blurry to most of its stakeholders, so this definitely needs to be improved' (B48).

In the DUTCHCO case many of the negative responds on communication where about bad communication between teams. Overall, a generic observation can be made on communication:

Observation 5: Satisfied stakeholders emphasize good communication. Dissatisfied stakeholders say communication needs to be improved.

4.5.3 Requirements (A4)

Most of the comments related to requirements were about unclear requirements that hinder a project's progress, such as:

'Interpretation from requirements can be different and cause issues at testing phase' (B40).

A limited number of comments were made on bad documentation, design problems and requirements creep, but also some comments were made on the availability of good requirements.

In the specific DUTCHCO case several remarks were made on unclear or hidden business rules as a cause for problems. Yet, a generic observation can be made with regard to requirements in both studied companies:

Observation 6: Dissatisfied stakeholders emphasize unclear requirements, bad documentation, hidden business rules, and requirements creep.

4.5.4 Stakeholder Satisfaction and Duration (A5, A6)

Many of the comments related to stakeholder aspects were about satisfied stakeholders. Most comments had to do with the quality of delivery and the time-to-market of delivery. Project duration and time-to-market was mentioned by many participants, where most comments are about on-time delivery. In the BELTEL case the following observation applied:

Observation 7: Satisfied stakeholders comment about good quality of duration estimates. Dissatisfied stakeholders comment about long duration and schedule overrun.

In the specific DUTCHCO case many respondents indicate they are satisfied with the product that was delivered to DUTCHCO's customer, as for example stated by D01:

'Despite the time constraints a decent product was delivered'.

With regard to this some respondents mentioned the inclusion of the customer as a positive factor, as stated by D10:

'I am satisfied about the fact that we included the customer and got their feedback, which resulted in a better product'.

However, with regard to *Project Duration* a number of DUTCHCO respondents mention a high time pressure, especially towards the end of the release, leading to issues and defects in the last stages. As D13 says it:

'Time issues caused several problems'.

Although duration was mentioned by stakeholders from DUTCHCO, they did this from the perspective of time pressure at the end of the release. In the DUTCHCO case no estimations were prepared with regard to duration. However, in the original BELTEL study we found that satisfied stakeholders comment about good quality of duration estimates, where dissatisfied stakeholders comment about long duration and schedule overrun. In the DUTCHCO study we found one additional observation with regard to time pressure:

Observation 8: Time pressure towards the end of a release leads to more defects.

Table 9. Results of the analysis of free format text of the BELTEL survey.

	Points of attention for satisfaction and value	Count
A1	Quality (good quality 27, bad quality 12)	39
A2	Communication (good communication 21, bad communication 17)	38
A3	Deployment (issues with implementation 19, issues in production 9, bad or delayed implementation 9)	37
A4	Requirements (requirements not clear 15, good requirements 9, requirements creep 5, bad documentation or design problems 4)	33
A5	Stakeholders (satisfied stakeholders 29, low stakeholder involvement 3, unsatisfied stakeholders 1)	33
A6	Duration (good estimation of duration; in-time delivery 16, bad estimation of duration 7)	23
A7	Testing (good testing or good test environment 8, problems with testing 9, delayed testing 3	20
A8	Process (smooth, lean, or mature process 11, (agile) process needs improvement 3, bad process 2, process not according to standards 2)	18
A9	Project Management (scope problems 10, good project management 3, scope delivered 3)	16
A10	Agile Development (good product owner 4, good backlog management tool 2, use of tools unclear 1, agile process needs improvement 1, traditional release in agile process 1)	14
A11	Supplier Management (issues with supplier 11, good relation with supplier 1, bad alignment between parties 1)	13
A12	Team Aspects (good team spirit 7, team not fixed 1)	8
A13	Release Management (bad alignment project and release 6, release delayed 1)	7
A14	Value Aspects (good value 2, issues with value 4)	6
A15	Cost Aspects (within time and budget)	1

Table is sorted on Count.

Table 10. Results of the analysis of free format text of the DUTCHCO survey.

	Points of attention for satisfaction and value	Count
A7	Testing (issues and defects late in the release 10, testing to be improved 8, improve acceptance criteria 3, large number of defects 2)	23
A14	Value Aspects (innovative solution, yet customer specific 15, many new features, good business value for the customer 7)	22
A4	Requirements (changes in scope and unclear features 9, hidden business rules as cause for problems 9, improve application 3)	21
A5	Stakeholders (satisfied stakeholders 3, customer driven 17)	20
A2	Communication (good communication between teams 5, to be improved communication between teams 3, better communication with customer 5)	16
A6	Duration (time pressure at the end of a release 11, strict timelines 2)	13
A8	Process (the internal process needs improvement, but people notice the company is changing for the good too 10, no proper process in place 2)	12
A1	Quality (satisfaction about the delivered results 8)	8
A12	Team Aspects (cooperation between teams, team performance, high pressure on team 7)	7
A9	Project Management (poor communication about scope of a release 5)	5
A15	Cost Aspects (profit made 1, paid by customer, yet generic 1)	2
A10	Agile Development	1
A3	Deployment	0
A11	Supplier Management	0
A13	Release Management	0

Table is sorted on Count.

4.5.5 Agile, Value, and Process (A10, A14, A8)

A more agile delivery process is one of the key innovations that are implemented within the software delivery organization of BELTEL, as well in the DUTCHCO case. Knowing this we argue that the low number of comments related to this aspect by BELTEL stakeholders (14) does not reflect the strategic choice of BELTEL for a new delivery approach, including the investments made in coaching and implementing tools that support an agile way or working. Eight (8) comments were positive about the quality of the product owner and the backlog management tool in use.

However, some comments were related to the agile process itself that needed improvement, as stated for example by B48:

'The whole agile process is still pretty blurry to most of its stakeholders so this definitely needs to be improved'.

For an organization that made delivery of value a strategic innovation remarkably few comments were made on value aspects. Two were about good value being delivered, while most had to do with the lack of value, such as:

'No real feeling on the benefit of this project' (B45).

With regard to process aspects a limited number of comments were about needs for improvement, such as speeding up things and working in a more structured way. However, about as many comments were related to a lean and flexible process.

In the specific DUTCHCO case we observed that from the viewpoint of innovation stakeholders overall seem to be quite happy with the delivered result, although some respondents mention the fact that the release was only applicable to one specific customer, as for example stated by D21:

'This was a big step in innovation, which would help us move further old applications to new technology stack' (D21).

Some stakeholders mention that not only the delivered product is innovative, but also the applied internal process:

'New technology used, new groundwork for new applications has been set. New way of working also (introduction of design street, and QA-department)' (D10).

Observation 9: Satisfied stakeholders emphasize the delivery of good value to the customer.

However strongly related to remarks on to be improved internal communication, stakeholders feel that DUTCHCO's internal process should be improved, although people see things changing for the good too, as for example mentioned by D01:

'The new structure of the company sort of has evolved on a better level'.

Although value certainly was addressed by DUTCHCO's stakeholders, no comments were made related to the agile delivery approach. Where in the specific BELTEL case an observation applied that the low number of comments related to agile processes does not reflect the company's strategic choice for a new delivery approach, we argue that the DUTCHCO omission is only partially comparable with the BELTEL study; DUTCHCO stakeholders simply do not talk about agile because it is their only delivery approach. Thus, we'd like to adjust the BELTEL observation to the following generic one:

Observation 10: 'Agile' itself is not always a point of discussion in companies, even when they are agile.

4.5.6 Supplier Management (A11)

A number of comments were about issues with suppliers, where also BELTEL's main supplier INDSUP was mentioned several times, such as for example by B14:

'Very long delays and complete lack of knowledge and initiatives from INDSUP'.

In the DUTCHCO case no specific comments were found that related to supplier management. In a way this is not surprising, since besides the INDSUB teams no external parties are applicable within DUTCHCO.

4.5.7 Cost

A remarkable observation is that only once a comment is made related to cost of projects in the BELTEL study:

'Implementation as per time, budget, and quality' (B38).

No comments were made about the estimation accuracy with regard to project cost. The aspect of cost was mentioned only twice by DUTCHCO stakeholders.

'We got paid quite a sum of money for this and I think we made some profit' (D10).

'Solution was completely paid by customer where we can use this functionality as generic functionality in our own solution' (D03).

Although the backgrounds may be different, like BELTEL cost seems not a large issue for DUTCHCO stakeholders:

Observation 11. Cost does not seem an important issue for project stakeholders.

5. DISCUSSION

In order to validate the outcome of the quantitative analysis with the outcome of the qualitative analysis, we compared the observations from both analyses, as depicted in Table 11. We grouped the data into four themes that correspond to the horizontal rows in Table 8 that logically belong together; the core project metrics (*Project Size*, *Project Cost*, *Project Duration*, and *Number of Defects*), *Stakeholder Satisfaction*, *Perceived Value*, and the quality of duration estimations.

5.1 The Core Project Metrics

The strong positive correlations that we found between *Project Size* on one hand and *Project Cost*, *Project Duration*, and *Number of Defects* (observation 1), confirm what is already known from related work [45] [5] [46] [47] [48] [49]. From this point of view, *Project Size*, measured in *function points*, can be considered as a very strong predictor of both cost and process quality.

Also the effect of project size as a risk factor is described earlier. Smaller projects tend to have lower cancellation rates [50] [51] [52]. Smaller projects tend to perform better in terms of quality, being on budget, and being on schedule [50] [51] [52] [53]. Project size is found to be an important risk factor for success [54] [55] [56] [57] [58] [59]. Note, however, that the literature does not match results from our study with regard to an economy-of-scale effect that larger projects in size are good, for Cost per FP and Days per FP (see paragraph 4.4). We argue that for most projects a trade-off is applicable between cost and duration on the one side, and risk on the other.

Despite the strong correlation, the use in practice of function points to measure project size suffers from shortcomings, such as additional training needed, subjective determination of complexity, and not considering the development environment [60].

In order to emphasize the effect of functional size (Function Points) as a normalizer we give an example related to the cheapest of all BELTEL projects versus the most expensive one. Project BelTel 6.3, a small (16 FPs) release-based enhancement on a CRM-application that was performed in a Scrum way as depicted above represents the minimum cost of 8,000 euro. This project scores good *Cost Duration Matrix* in Figure 2, and shows the highest score of all for *Stakeholder Satisfaction* for both process and result. To put things in perspective, the maximum cost of 296,000 euro is linked to Project BelTel 6.10, an implementation of a part of a new order management system. Yet, also this project scores on the upper side in the *Cost over Time* quadrant, mainly due to a high number of function points that are delivered; 324 FPs. This project also scores well for both *Stakeholder Satisfaction* and *Perceived Value*. We note that both projects were performed in a Scrum way as depicted above.

With regard to the strong correlations that we found between Project Size and other core project metrics such as cost and defects, it might be important to consider that Project Size was measured in Function Points, in a manual counting process. As described in the data collection approach, different artifacts were used as a source for function point counting, depending on the availability per project (e.g. sets of functional documentation, user stories recorded in one of the Scrum backlog tools, architectural documents, project documentation, user manuals, or wireframes). Manual counting was performed by different members of measurement teams of both companies, and reviewed by another member, to ensure proper use and interpretation of counting guidelines [30].

In agile environments, where usually no upfront artifacts such as functional and technical design documents are prepared, counting functional size can be challenging. However, we experienced in practice in both companies that

Table 11. Summary of observations and implications for practice and research.

Themes	Observations from the quantitative analysis	Observations from the qualitative analysis	Implications for practice and research
The Core Project Metrics	Strong positive correlations are found between Project Size, Project Cost, and Number of Defects (observation 1).	Cost does not seem an important issue for project stakeholders (observation 11). Time pressure towards the end of a release leads to more defects (observation 8) (DUTCHCO).	The strong correlations between Project Size on the one hand, and Project Cost, Project Duration, and Number of Defects on the other are confirmed by related work [44] [5] [46] [47] [48] [49].
Stakeholder Satisfaction	Stakeholder Satisfaction for both process and result are strongly interrelated to each other. (observation 2). Stakeholder Satisfaction relates negatively with Project Duration (observation 2).	Satisfied stakeholders emphasize good quality (observation 4), good communication (observation 5), and good quality of duration estimations (observation 7). Dissatisfied stakeholders state that testing and deployment needs improvement (observation 4), they emphasize unclear requirements, bad documentation, hidden business rules, and requirements creep (observation 6), they say communication needs to be improved (observation 5), and they comment about long duration and schedule overrun (observation 7).	Additional research is needed to identify how good quality of deliverables, good communication, and reliable estimations for duration can be controlled in practice in order to create satisfied stakeholders. Additional research is needed to identify how issues with testing and deployment, unclear requirements, hidden business rules, bad communication, and schedule overrun can be controlled in order to mitigate stakeholder dissatisfaction.
Perceived Value	Although the different Perceived Value measures interrelate strongly with each other, we cannot draw general conclusions from this. Weak correlations are found between Perceived Value and Project Size, Cost per FP, and Days per FP, we cannot draw general conclusions from this BELTEL phenomenon.	Satisfied stakeholders emphasize the delivery of good value to the customer (observation 9) (DUTCHCO). 'Agile' itself is not always a point of discussion in companies, even when they are agile (observation 10).	Additional research is needed (especially on medium and big sized projects) to validate whether larger projects tend to lead to higher perceived value scores. If this effect is true, functional size might be a potential indicator for value, that can be used in practice by Product Owners to prioritize (enterprise) backlogs.
Estimation Quality for Project Duration	Weak correlations are found between Estimation Quality Factor for Duration and Project Duration, and Stakeholder Satisfaction (observation 3).	Satisfied stakeholders emphasize good quality of duration estimations (observation 7). Dissatisfied stakeholders comment about long duration and schedule overrun (observation 7). Time pressure towards the end of a release leads to more defects (observation 8) (DUTCHCO).	Much research is performed on Effort Estimation of software projects [72], yet very limited research is performed on the effects of the quality of Duration Estimation, because the outcomes of this study indicate correlations with both stakeholder satisfaction and perceived value.

descriptions of user stories in backlog tools together with additional information such as wireframes, where suitable to perform a reliable function point count. In all cases so-called estimated Function Point counts were performed.

In order to automate the data collection process where possible, we strongly felt a need for some form of automated function point count, if possible based on the code itself. However, a follow-up exploratory study of 336 functional size measurement specialists that was performed based on this hypothesis did showed that overall automated functional size measurement was considered as important, but also difficult to realize [61].

5.2 Stakeholder Satisfaction

Observation 2, indicates a moderate correlation between *Stakeholder Satisfaction* for both process and result and *Project Duration*. We found a moderate relation between *Stakeholder Satisfaction* for *Estimation Quality Factor (Duration)* too. This indicates that *Stakeholder Satisfaction* is related to interaction and being informed, yet also with conformance to planning and estimation of the delivery date. A strategy of 'no last minute surprises' as such helps better to increase *Stakeholder Satisfaction*, as well as giving attention to improvement of estimation and planning practices would.

Satisfaction of stakeholders with the development process and with the development outcome was studied before by Ferreira and Cohen [62]. They found "strong positive effects of agile practice (iterative development, continuous integration, collective ownership, test-driven design, and feedback) on stakeholder satisfaction with both development process and the project outcome". A relation between stakeholder satisfaction and with agile software development in an Indian context was found by Nazir et al. [63].

Many participants mentioned communication to be important, while good communication is mentioned by satisfied stakeholders, and bad communication by dissatisfied ones. Approximately half of the comments were about good communication, such as good alignment between parties, good collaboration, and short feedback loops. The other half mention communication to be improved, such as provide information on processes and innovations (e.g. agile delivery), ongoing discussions, and miscommunication with suppliers. Unclear requirements, bad documentation, requirements creep, and bad quality of test and deployment resources are perceived by dissatisfied stakeholders as causes for bad quality of deliverables.

However, a warning is in place here: we notice that many positive comments on communication within BELTEL also are linked to two specific Product Owners. We did not focus our research on roles within the subject projects, but this suggests that the fulfillment of a role by a specific person may be of greater influence on *Stakeholder Satisfaction* and *Perceived Value* than the subject delivery model. Note that this resonates with the first line of the Agile Manifesto: "Individuals and Interactions over Processes and Tools" [64].

Contrary to the findings in our original study, we did not find evidence for correlations of *Stakeholder Satisfaction* with *Number of Defects*. However, in the qualitative analysis we do find many comments that are in one way or another related to those aspects. Quality of the deliverables (both good quality and to be improved quality), in combination with testing aspects and deployment into the production environment, is commonly mentioned in comments by all participants.

5.3 Perceived Value

We assume that the relative absence of comments that are related to the ongoing innovation of implementing a more agile delivery process within BELTEL, in combination with the limited focus on value might be of importance here (observation 7). The low interest in agile innovation among BELTEL's stakeholders in a way reflects our findings in the quantitative analysis too. No significant relation is found between *Perceived Value* and any other project metric of software project deliverables. However, the limited number of projects in scope of this study, combined with the diversity in project sizes (with many small and only one large project) can be a reason for replicating our study with more data in future.

Furthermore, what strikes with regard to the value measures are the almost perfect correlation coefficients (from 0.94 to 0.99) with p-value < 0.001 that we found between all four *Perceived Value* measures mutually. This is not a good sign when designing scales. It might imply that the four aspects, derived from well-known research on Kaplan and Norton's Balanced Scorecard [42], are measuring the same construct, or that the answers to those items were influenced in the same way. Besides that a warning is in place here with regard to functional relations between the perceived Value metrics. As it is dubious practice to correlate metrics that have a functional relationship between them, as likely spurious correlations are found [44], we do not valid the findings on interrelated Perceive Value metrics as reliable.

5.4 Estimation Quality for Duration

The comments given in the surveys confirm the more or less company specific observations with regard to *Project Duration*. Stakeholders of projects are satisfied when delivery of results is in-time, where we assume this relates to good quality of duration estimates. However, it needs to be said that the words *estimate* or *estimation* are never used in the comments. Dissatisfaction of stakeholders is often linked with too late delivery and long project durations (long waiting time).

In the DUTCHCO case we do not find direct evidence for this, although time pressure towards the end of a release, causing issues and defects at a late stage in the release, are mentioned as a source for problems by many respondents.

5.5 Success or failure: complex relations

Looking at the seventh row and eight column in Table 8, the test results of association between paired samples, it strikes that no correlation is found at all between *Cost Duration Index* and any of the other samples. Apparently no relation exists between the position of a project in our *Cost Duration Matrix* and the measure of Stakeholder Satisfaction and perceived Value of that project. Based on this we conclude that success and failure apparently are more complex than cost, duration and defects only: stakeholders can be satisfied or have the perception of much value delivered, even when a project is in the so-called Bad Practice quadrant.

We suspect, based partly on recent ongoing research, that the limitation in the current study to internal stakeholders of projects can play a limiting role here. For external stakeholders, usually the customers of the software organizations that actually pay for the software and use it in practice, cost, duration and number of errors seem to be a very important factor for success or failure. Besides that we suspect that Perceived Value needs to be measured on a lower level than a project (e.g. at user story or at epic level). We argue that additional research is needed to unfold these complex relations.

5.6 Agile and Cost were not mentioned

Except for the last four themes that resulted from the quantitative and qualitative analyses, we found two issues that were not mentioned in the text free format of the surveys, and that were not found in the tests for association between paired samples.

Firstly, we found a very low number of comments in both studies, that are related to the concept of 'agile' itself. A number of observations that were applicable for BELTEL did not apply to DUTCHCO, and the other way around. No evidence was found within DUTCHCO on the relation between satisfied stakeholders and good quality of estimates. DUTCHCO does not produce estimates as such for its software delivery activities. Like BELTEL, within DUTCHCO no specific remarks were made about the agility of its process. However, an agile development approach within DUTCHCO is widely assumed as the only form of process, no other approaches (e.g. plan-driven) are applicable. Thus, we assume that no comments were made about this just because it is 'business as usual'.

Besides the development method used, we assume that also differences in test and deployment (e.g. release)

approaches of both software companies influence the portfolio performances. The fact that DUTCHCO collects all of its features in large three-monthly releases, while BELTEL runs a variety of software projects sequentially with eight combined releases per year, should be taken into account when comparing the performances of both companies.

A second finding was the fact that cost seems not an important issue for stakeholders within both BELTEL and DUTCHCO. Only one comment by a BELTEL stakeholder is made related to this. This finding applies to DUTCHCO too; not much attention is given in the comments on cost either. We think that this might be caused by the effect that in more or less agile organizations, the focus tends to shift from time and cost driven controlling towards scope and value driven steering. Agile teams tend to stay in place for longer periods, and budgeting often needs to be done only once a year, instead of many times per year in a pre-project phase in plan-driven organizations.

In a way this is not a big surprise in the DUTCHCO organization, since development teams are fixed and stay together for long periods of time. Due to this cost is just a derivative of effort spent by these teams. No budget estimations upfront are applicable in the DUTCHCO organization. Although this cannot be understood in a way that DUTCHCO is not interested in cost at all; DUTCHCO's management team is highly interested in cost reduction based on the effects of shortening learning curves [28] and efficiency improvements based on outsourcing approaches.

5.7 Implications

The outcomes of both our case studies might not simply be generalized to other environments. We identify a number of take-away-messages that apply to research and practice in other software companies too.

The first one relates to the from related work already known strong correlations between *Project Size* on the one hand, and *Project Cost*, *Project Duration*, and *Number of Defects* on the other, indicating the power of *Project Size* (measured in Function Points) as an indicator for cost, duration, and quality. However, a link with agile development is poorly covered in research [61], and in practice many agile software companies tend to see (manual) counting of functional size as waste. We argue that also agile practitioners and researchers should rethink and embrace *Project Size*.

Secondly, we argue on the one hand that good quality of deliverables, good communication, and reliable estimations for duration can be used to increase stakeholder satisfaction, and on the other that issues with testing and deployment, unclear requirements, hidden business rules, bad communication, and schedule overrun increase stakeholder dissatisfaction, and should therefore be mitigated when possible. Assuming that agile development methods might play a role here, additional research is needed to identify the backgrounds and ways to control these findings in a practical context.

Thirdly, we recognize a need for additional research (especially on medium and large sized projects) to validate whether larger projects tend to lead to higher perceived value scores. We think that measuring perceived value on a lower level than a software project, e.g. on user stories or epics, might result in other outcomes. If a strong positive effect is found in future, functional size might be a potential indicator for value, that can be used in practice by Product Owners to prioritize (enterprise) backlogs.

Finally, a fourth take-away-message relates to a need for additional research in a practical context on the effects of the quality of Duration Estimation, because the outcomes of this study indicate correlations with both stakeholder satisfaction and perceived value. Good quality estimation of a project's delivery date seems very important for stakeholders, and relates to the perception of value that is delivered.

5.8 Threats to Validity

5.8.1 Construct Validity

With regard to construct validity constraints we emphasize that we asked stakeholders for perceptions on satisfaction and value. Perceptions are not the same as actual measurements, which is especially the case for our value measurements. We prefer to measure the real business value as delivered by each software project. However, two problems occur with regard to this. Holistic measurements on value are often difficult to make for a single project (e.g. *Return on Investment* and *Net Present Value*). Besides that, such measures (e.g. *Net Promotor Score*) cannot easily be related to software projects, mainly because too many different factors are of influence for such measurements.

As explained in paragraph 5.3, two limitations are in place with regard to the setup of our electronic survey. Adopting any of the existing validated measurement instruments on customer / stakeholder satisfaction, which are available from the behavioral science, and economics and management theory, might be helpful for continuation of the survey in future research. Secondly, the almost perfect correlations between the four aspects of Perceived Value indicate that the aspects are measuring the same construct, or that the answers to those items were influenced in the same way. We argue that it would be good to adjust the survey with regard to these aspects for future research.

5.8.2 Internal Validity

A threat to internal validity that we acknowledge is the fact that 'fishing for p-values' might hold a risk that some of the correlations we find are a coincidence. We limited this effect by making Benjamini-Hochberg corrections for all p-values that we used in the multiple comparisons (see Table 8). Furthermore, the number of parameters in our model is too low to perform a reliable generalized linear model test with multiple data points. To prevent from systematic error we perform an exploratory test in which we do test for p-values, yet we confront these with findings of the qualitative analysis.

In order to minimize systematic error with regard to subjectiveness of stakeholders in their survey answers, we included representatives from both IT and business that were involved in any way in a subject project. We considered to also include participants (from the organizations involved) that did not know the subject projects in the assessment of perceived value. However, the study was performed in an operational context within BELTEL and subsequently DUTCHCO.

Answering surveys, subsequent a release, was implemented as an operational capability. When designing the study we considered that it was undesirable to interfere with stakeholders more than necessary in their operational activities, and not to engage them in surveys related to projects in which they did not participate.

Another attempt we made to prevent from bias, was to perform anonymous surveys, although one can argue that based on specific roles a lack of anonymity could introduce potential bias. In order to reduce bias due to ambiguity of survey answers with regard to the four aspects of value (customer, internal process, financial, and innovation) we applied additional text on the survey that was shown when participants hovered over a question mark linked to each question.

A limitation is in place regarding the summarized themes in Table 9 and Table 10; the included key concepts are defined loosely on free text concepts provided by respondents in which they later were categorized.

5.8.3 External Validity

Concerning external validity, the extent to which the results of our study can be generalized to other companies than BELTEL and DUTCHCO is difficult to answer because we performed multiple case studies in these two specific companies. Not all findings that occurred in the BELTEL case were found in the DUTCHCO case too. Especially because our findings relate to specific situations, maturity, and development approaches we argue that a one-on-one generalization to other companies is not valid. Instead we argue that evidence-based software engineering [11] in a way we performed for this study within both BELTEL and DUTCHCO is a precondition for mature improvement within other companies too.

5.8.4 Study Reliability

A threat related to the study's reliability lies in the fact that the lead author of this paper was a member of the measurement team within BELTEL, and carried out the functional size measurements within DUTCHCO. However, we tried to prevent from bias by ensuring that the BELTEL measurement team and DUTCHCO measurement expert are independent and objective in their collection of data. Moreover, the size measurements were made before the analyses were made and performed along the functional size measurement procedures that are repeatable independently from the actual measurement expert [30].

A link that we did not study, but that is mentioned in other studies, is the relation between stakeholder satisfaction and requirements [65] [66] and documentation [67].

6. CONCLUSIONS AND FUTURE RESEARCH

The outcomes of our multiple case study indicate, that "within time and cost" does not automatically lead to satisfied stakeholders. A focus on shortening overall project duration, and good communication (e.g. no last minute surprises) and optimal collaboration between teams, has a positive effect on satisfaction of stakeholders. On the other hand, too late delivery and long project durations, and many defects dissatisfy them. Our study does not provide any evidence that steering on costs helped to improve the satisfaction of stakeholders.

A novelty in the results of our study is that we linked *Perceived Value* to a set of project metrics, among others functional size of projects. As an answer to our research question - *how do stakeholder satisfaction and perceived value relate to cost, duration, defects, size and estimation accuracy of software projects?* – we found the following five take-away-messages:

- *Stakeholder Satisfaction* can be improved by steering on good quality, good communication, and good quality of duration estimations. Satisfied stakeholders emphasize the delivery of good value to the customer.
- *Stakeholder Satisfaction* goes down when issues occur with testing and deployment, unclear requirements, bad documentation, hidden business rules, and requirements creep, when communication is bad, and in case of long project duration and schedule overrun.
- *Perceived Value* correlates (however weakly) with *Project Size*, *Cost per FP*, and *Days per FP*, indicating that functional size might be an indicator for value, however more research is needed to confirm this finding.
- We identified two themes that did not apply to stakeholders or value: 'Agile' itself is not always a point of discussion in companies, even when they are agile, and *Project Cost* seems not an important issue for stakeholders.
- The study also confirmed an effect known from related work: *Project Size* strongly correlates with the other core project metrics *Project Cost*, *Project Duration*, and *Number of Defects*.

Our final question is how we and others build on the main findings of this study. We see the following four aspects to be important for further research:

- How can good quality of deliverables, good communication, and reliable estimations for duration be controlled to create satisfied stakeholders?
- How can issues with testing and deployment, unclear requirements, hidden business rules, bad communication, and schedule overrun be controlled to mitigate stakeholder dissatisfaction?
- Do larger projects (in functional size) lead to higher perceived value scores?
- How can quality of project duration estimations be used to improve stakeholder satisfaction and perceived value?

ACKNOWLEDGMENTS

We thank BELTEL, DUTCHCO, and INVEND for their generosity to allow us to use company data in our research, and all survey respondents for their help on sharing their ideas on improvement of software projects with us. We thank Tableau for allowing us to use their BI solution to build our *Performance Dashboard*.

REFERENCES

[1] International Standish Group, "The Chaos Report," 1994. [Online]. Available: www.standishgroup.com.

[2] M. Jørgensen and K. Moløkken-Østvold, "How large are software cost overruns? A review of the 1994 CHAOS report," *Information and Software Technology*, vol. 48, no. 4, pp. 297-301, 2006.

[3] R. L. Glass, "The Standish Report: Does It Really Describe a Software Crisis?," *Communications of the ACM*, vol. 49, no. 8, pp. 15-16, 2006.

[4] J. L. Eveleens and C. Verhoef, "The rise and fall of the chaos report figures," *IEEE software*, vol. 27, no. 1, pp. 30-36, 2010.

[5] H. Huijgens, R. van Solingen and A. van Deursen, "How to build a good practice software project portfolio?," in *ACM Companion Proceedings of the 36th International Conference on Software Engineering (ICSE)*, 2014.

[6] H. Huijgens, A. van Deursen and R. van Solingen, "An Exploratory Study on the Effects of Perceived Value and Stakeholder Satisfaction on Software Projects," in *Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, Limerick, Ireland, 2016.

[7] H. Huijgens, G. Gousios and A. van Deursen, "Pricing via Functional Size - A Case Study of a Company's Portfolio of 77 Outsourced Projects," in *ACM/IEEE 9th International Symposium on Empirical Software Engineering and Measurement (ESEM)*, Beijing, China, 2015.

[8] T. Gilb and S. Finzi, Principles of software engineering management, vol. 11, Reading, MA: Addison-Wesley, 1988.

[9] K. Moløkken and M. Jørgensen, "A Review of Surveys on Software Effort Estimation," *IEEE - Proceedings of ISESE - International Symposium on Empirical Software Engineering*, pp. 223-230, 2003.

[10] K. El Emam and A. Gunes Koru, "A replicated survey of IT software project failures," *IEEE Software*, vol. 25, no. 5, pp. 84-90, 2008.

[11] T. Dyba, B. A. Kitchenham and M. Jorgensen, "Evidence-based software engineering for practitioners," *Software, IEEE*, vol. 22, no. 1, pp. 58-65, 2005.

[12] R. L. Glass, "IT Failure Rates-70% or 10-15%?.," *IEEE Software*, vol. 22, no. 3, pp. 110-112, 2005.

[13] J. Sutherland, A. Viktorov, J. Blount and N. Puntikov, "Distributed Scrum: Agile Project Management with Outsourced Development Teams," in *40th International Conference on System Sciences*, Hawaii, 2007.

[14] J. Pekki, "How the Company Manages Critical Success Factors in Software Process Improvement Initiatives: Pilot Case-Study in Finnish Software Company." European Conference on Software Process Improvement. Springer International Publishing, 2016.," in *European Conference on Software Process Improvement*, 2016.

[15] K. Beck, Extreme Programming Explained: Embrace Change, Addison-Wesley, 2000.

[16] T. Dingsøyr and C. Lassenius, "Emerging themes in agile software development: Introduction to the special section on continuous value delivery," *Information and Software Technology*, vol. 77, pp. 56-60, 2016.

[17] R. Atkinson, "Project management: cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteria," *International journal of project management*, vol. 17, no. 6, pp. 337-342, 1999.

[18] M. Shepperd, "Cost prediction and software project management," in *Software Project Management in a Changing World*, Springer, 2014, p. 51–71.

[19] K. Strand and K. Karlsen, "Agile Contracting and Execution," in *PROMIS*, 2014.

[20] H.-M. Chen, R. Kazman, J. Garbajosa and E. Gonzalez, "Toward big data value engineering for innovation," in *ACM Proceedings of the 2nd International Workshop on BIG Data Software Engineering*, 2016.

[21] B. Boehm, "Value-Based Software Engineering," *ACM SIGSOFT Software Engineering Notes*, vol. 28, no. 2, pp. 1-12, 2003.

[22] S. Biffl, A. Aurum, B. Boehm, H. Erdogmus and P. Grünbacher, Value-Based Software Engineering, Berlin Heidelberg: Springer, 2006.

[23] S. Faulk, D. Harmon and D. Raffo, "Value-Based Software Engineering (VBSE): A Value-Driven Approach to Product-Line Engineering," in *First International Conference on Software Product-Line Engineering*, 2000.

[24] N. Agarwal and U. Rathod, "Defining 'success' for software projects: An exploratory revelation," *International journal of project management*, vol. 24, no. 4, pp. 358-370, 2006.

[25] D. J. Bryde, "Methods for managing different perspectives of project success," *British Journal of Management*, vol. 16, no. 2, pp. 119-131, 2005.

[26] M. Jørgensen, "A survey on the characteristics of projects with success in delivering client benefits," *Information and Software Technology*, vol. 78, pp. 83-94, 2016.

[27] K. Schwaber and J. Sutherland, "The scrum guide," Scrum Alliance 21, 2011. [Online]. Available: www.scrumguides.org.

[28] D. Smite and R. v. Solingen, "What's the True Hourly Cost of Offshoring?," *IEEE Software*, vol. 33, no. 5, pp. 60-70, Sept-Oct 2016.

[29] H. Huijgens, "Evidence-based software portfolio management: a tool description and evaluation," in *ACM Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, 2016.

[30] IFPUG, IFPUG FSM Method: ISO/IEC 20926 - Software and systems engineering – Software measurement – IFPUG functional size measurement method, New York: International Function Point User Group (IFPUG), 2009.

[31] T. DeMarco, "An algorithm for sizing software products," *ACM SIGMETRICS Performance Evaluation Review*, vol. 12, no. 2, pp. 13-22, 1984.

[32] L. J. Eveleens and C. Verhoef, "Quantifying IT forecast quality," *Science of computer programming*, vol. 74, no. 11, pp. 934-988, 2009.

[33] C. Wu, M. Naqibuddin and L. Fleisher, "Measurement of patient satisfaction as an outcome of regional anesthesia and analgesia:

A systematic review," *Reg Anesth Pain Med*, vol. 26, pp. 196-208, 2001.

[34] G. Pascoe, "Patient satisfaction in primary health care: A literature review and analysis," *Eval Prog Planning*, vol. 6, pp. 185-201, 1983.

[35] P. Auquier, N. Pernoud, N. Bruder , M. Simeoni, J. Auffray, C. Colavolpe and C. Sapin, "Development and Validation of a Perioperative Satisfaction Questionnaire," *Clinical Science*, vol. Anesthesiology 6, no. 102, pp. 1116-1123, 2005.

[36] M. J. Atkinson, A. Sinha , S. L. Hass, S. S. Colman, R. N. Kumar, M. Brod and C. R. Rowland, "Validation of a general measure of treatment satisfaction, the Treatment Satisfaction Questionnaire for Medication (TSQM), using a national panel study of chronic disease," *Health and Quality of Life Outcomes*, vol. 2, no. 12, 2004.

[37] B. E. Hayes, Measuring customer satisfaction: Survey design, use, and statistical analysis methods, ASQ Quality Press, 1998.

[38] R. van Solingen, "Measuring the ROI of software process improvement," *IEEE Software*, vol. 21, no. 3, pp. 32-38, 2004.

[39] P. Green, "Measuring the impact of scrum on product development at adobe systems," in *IEEE 44th Hawaii International Conference on System Sciences (HICSS)*, 2011.

[40] G. Hofner, V. Mani, R. Nambiar and M. Apte, "Fostering a high-performance culture in offshore software engineering teams using balanced scorecards and project scorecards," in *IEEE Sixth International Conference on Global Software Engineering*, 2011.

[41] M. Feyh and K. Petersen, "Lean software development measures and indicators-a systematic mapping study," *Lean Enterprise Software and Systems*, pp. 32-47, 2013.

[42] R. S. Kaplan and D. P. Norton, "Putting the balanced scorecard to work," *Harvard Business Review*, vol. 71, no. 5, pp. 134-140, 1993.

[43] W. G. Hopkins, "A new view of statistics," *Internet Society for Sport Science*, 2000.

[44] M. Jørgensen, T. Halkjelsvik and B. Kitchenham, "How does project size affect cost estimation error? Statistical artifacts and methodological challenges," *International Journal of Project Management*, vol. 30, pp. 839-849, 2012.

[45] B. Boehm, "Software Engineering Economics," *IEEE Transactions on Software Engineering*, vol. 10, no. 1, pp. 7-19, 1984.

[46] K. El Emam and A. Günes Koru, "A replicated survey of IT software project failures," *IEEE software*, vol. 25, no. 5, pp. 84-90, 2008.

[47] B. Boehm, C. Abts and S. Chulani, "Software development cost estimation approaches - A Survey," *Annals of Software Engineering*, vol. 10, pp. 177-205, 2000.

[48] F. Heemstra and R. Kusters, "Function point analysis: Evaluation of a software cost estimation model," *European Journal of Information Systems*, vol. 1, no. 4, pp. 223-237, 1991.

[49] M. Bhardwaj and A. Rana, "Key Software Metrics and its Impact on each other for Software Development Projects," *ACM SIGSOFT Software Engineering Notes*, vol. 41, no. 1, pp. 1-4, 2016.

[50] D. Rubinstein, "Standish group report: There's less development chaos today," *Software Development Times*, vol. 1, 2007.

[51] C. Jones, "Project Management Tools and Software Failures and Successes," *Crosstalk*, vol. July, pp. 13-17, 1998.

[52] C. Sauer and C. Cuthbertson, "The State of IT Project Management in the UK 2002–2003," *Computer Weekly*, vol. April, 2003.

[53] R. Sonnekus and L. Labuschagne, "Establishing the Relationship between IT Project Management Maturity and IT Project Success in a South African Context," Proc. 2004," *PMSA Global Knowledge Conf., Project Management South Africa*, pp. 183-192, 2004.

[54] H. Barki, S. Rivard and J. Talbot, "Toward an assessment of software development risk," *Journal of Management Information Systems*, vol. 10, pp. 203-223, 1993.

[55] J. Jiang and G. Klein, "Software development risks to project effectiveness," *Journal of Systems and Software*, vol. 52, no. 1, pp. 3-10, 2000.

[56] R. Schmidt, K. Lyytinen, P. Cule and M. Keil, "Identifying software project risks: An international Delphi study," *Journal of management information systems*, vol. 17, no. 4, pp. 5-36, 2001.

[57] D. Zowghi and N. Nurmuliani., "A study of the impact of requirements volatility on software project performance," in *IEEE Ninth Asia-Pacific Software Engineering Conference,*, 2002.

[58] F. Heemstra and R. Kusters, "Controlling Software Development Costs: A Field Study," in *International Conference on Organisation and Information Systems*, Bled, Yugoslavia, 1989.

[59] R. B. Chidambara and M. Senthil Kumar, "Contribution of risk assessment techniques in effort estimation of software development process," *PARIPEX-Indian Journal of Research*, vol. 5, no. 4, 2016.

[60] B. Singh, A. Punhani and A. Misra, "Integrated Approach of Software Project Size Estimation," *International Journal of Software Engineering and Its Applications*, vol. 10, no. 2, pp. 45-64, 2016.

[61] H. Huijgens, M. Bruntink, T. van der Storm, F. Vogelezang, and A. van Deursen, "An exploratory study on functional size measurement based on code," in *ACM Proceedings of the International Conference on Software and Systems Process (ICSSP)*, 2016.

[62] C. Ferreira and J. Cohen, "Agile systems development and stakeholder satisfaction: a South African empirical study," in *ACM Proceedings of the 2008 annual research conference of the South African Institute of Computer Scientists and Information Technologists on IT research in developing countries*, 2008.

[63] N. Nazir, N. Hasteer and A. Bansal, "A survey on agile practices in the Indian IT industry," in *IEEE 6th International*

Conference-Cloud System and Big Data Engineering (Confluence), 2016.

[64] K. Beck, M. Beedle, A. v. Bennekum, A. Cockburn, W. Cunningham, M. Fowler and J. Kern, "Manifesto for Agile Software Development," 2012. [Online]. Available: www.agilemanifesto.org.

[65] A. M. Pitangueira, P. Tonella, A. Susi, R. Maciel and M. Barros, "Risk-Aware Multi-stakeholder Next Release Planning Using Multi-objective Optimization," in *International Working Conference on Requirements Engineering: Foundation for Software Quality*, 2016.

[66] R. S. Maciel and M. Barros, "Risk-Aware Multi-stakeholder Next Release Planning Using Multi-objective Optimization," in *22nd International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ)*, 2016.

[67] J. A. Díaz-Pace, C. Villavicencio , S. Schiaffino and M. Nicoletti, "Producing Just Enough Documentation: An Optimization Approach Applied to the Software Architecture Domain," *Journal on Data Semantics*, vol. 5, no. 1, pp. 37-53, 2016.

[68] J. H. McDonald, "Handbook of Biological Statistics," [Online]. Available: http://www.biostathandbook.com/multiplecomparisons.html. [Accessed 11 2016].

[69] A. Arcuri and L. Briand, "A hitchhiker's guide to statistical tests for assessing randomized algorithms in software engineering," *Software Testing, Verification and Reliability*, vol. 24, no. 3, pp. 219-250, 2014.

[70] Y. Hochberg, "A sharper Bonferroni procedure for multiple tests of significance," *Biometrika*, vol. 75, no. 800-2, 1988.

[71] B. Kitchenham, L. Madeyski, P. Brereton, S. Charters, S. Gibbs, A. Pohthong, D. Budgen and J. Keung, "Robust Statistical Methods for Empirical Software Engineering," *Empirical Software Engineering (in Press)*, 2016.

[72] M. Jørgensen and M. Shepperd, "A Systematic Review of Software Development Cost Estimation Studies," *IEEE Transactions on Software Engineering*, vol. 33, no. 1, pp. 33-53, 2007.

[73] H. Huijgens, A. van Deursen and R. van Solingen, "TUD-SERG-2017-001 - The Effects of Perceived Value and Stakeholder Satisfaction on Software Project Impact - Technical Report," Delft University of Technology, 2017.

7. Appendix A

7.1 R-script used for this research paper

```
setwd("C:/Users/Hennie/OneDrive/Documents/Promotieonderzoek/Elsevier_IST")
Repository <- read.csv("MeasurementRepository.csv",header=TRUE,sep=",",dec=".",fill=TRUE)
EBSPM <- read.csv("MeasurementRepositoryAll.csv",header=TRUE,sep=",",dec=".",fill=TRUE)

RepositoryBeltel = subset(Repository, Repository$Organization=='Beltel')
RepositoryDutchco = subset(Repository, Repository$Organization=='DutchCo')
RepositoryBeltelAll = subset(EBSPM, EBSPM$Organization=='Beltel')
RepositoryDutchcoAll = subset(EBSPM, EBSPM$Organization=='DutchCo')

summary(RepositoryBeltel)
summary(EBSPM)
summary(RepositoryDutchco)
str(RepositoryDutchco)
str(RepositoryBeltel)
str(RepositoryBeltelAll)
str(RepositoryDutchcoAll)
str(EBSPM)

cor.test(RepositoryBeltel$Cost, RepositoryBeltel$Size,
         use='pairwise', alternative = c("two.sided", "less", "greater"),
         method = c("spearman"),
         exact = NULL, conf.level = 0.95, continuity = FALSE)

cor.test(RepositoryBeltelAll$Duration, RepositoryBeltelAll$Size,
         use='pairwise', alternative = c("two.sided", "less", "greater"),
         method = c("spearman"),
         exact = NULL, conf.level = 0.95, continuity = FALSE)

cor.test(RepositoryDutchcoAll$Duration, RepositoryDutchcoAll$Size,
         use='pairwise', alternative = c("two.sided", "less", "greater"),
         method = c("spearman"),
         exact = NULL, conf.level = 0.95, continuity = FALSE)

# Statistical tests for comparing overall differences

wilcox.test(EBSPM$Size,Repository$Size)
wilcox.test(EBSPM$Cost,Repository$Cost)
wilcox.test(EBSPM$Duration,Repository$Duration)
wilcox.test(EBSPM$Defects,Repository$Defects)

# Statistical tests for comparing differences per size

wilcox.test(EBSPM$Cost/EBSPM$Size,Repository$Cost/Repository$Size)
wilcox.test(EBSPM$Duration/EBSPM$Size,Repository$Duration/Repository$Size)
wilcox.test(EBSPM$Defects/EBSPM$Size,Repository$Defects/Repository$Size)
```

```
install.packages("e1071")
library(e1071)
skewness(Repository$CostDurationIndex)
kurtosis(Repository$CostDurationIndex)
sd(Repository$CostDurationIndex)

hist(Repository$Cost)
hist(Repository$Size)
plot(Repository$Size, Repository$Cost)

install.packages("ggplot2")
library(ggplot2)
qplot(Repository$Size, Repository$Duration,
      colour = Repository$Organization, data = Repository,
      xlab = "Project Size", ylab = "Project Duration")

plot(Repository$Size, Repository$Duration,
     pch = c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,16,16,16,16),
     xlab = "Project Size", ylab = "Project Duration")

plot(RepositoryBeltelAll$Size, RepositoryBeltelAll$Duration)
plot(RepositoryDutchcoAll$Size, RepositoryDutchcoAll$Duration)

plot(Repository$Size, Repository$StakeholderSatisfactionProcess)
plot(Repository$Size, Repository$Stakeholder.SatisfactionProduct)
plot(Repository$Size, Repository$PerceivedValueOverall,
     pch = c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,16,16,16,16),
     xlab = "Project Size", ylab = "Perceived Value Overall")
shapiro.test(Repository$Cost)
shapiro.test(Repository$Size)

cor.test(Repository$PerceivedValueInnovation,    Repository$PerceivedValueOverall,    method    =
c("pearson"))

RepReg = lm(Size ~ Cost + Duration + Defects + PerceivedValue, data=Repository)
summary(RepReg)
```

7.2 Results of the statistical analysis

7.2.1 Summary of the BelTel subset

```
> summary(RepositoryBeltel)
       ID        Organization       ProjectID
 Min.   :390.0   Beltel:22    Project 3.1: 1
 1st Qu.:424.2   DutchCo: 0   Project 3.2: 1
 Median :429.5                Project 3.3: 1
 Mean   :426.4                Project 3.4: 1
 3rd Qu.:436.5                Project 3.5: 1
 Max.   :444.0                Project 4.1: 1
                              (Other)    :16

ProjectDescription
 Adapt T&C Procedure DOB Wmode Platform. Small enhancement to an existing system. Plan-driven de
livery approach. Performed by a team of the strategic partner (Indian, partly onsite, partly off
```

```
site). Project manager of the Telecom company (onsite).
: 1
 Admin Fee for SoHo Customers. Rules- and regulations driven small enhancement to existing CRM s
ystem. Small project team of onsite Project Manager and Business Analyst and offsite Indian deve
lopers.
: 1
 Apple Store SSO and QR Code. New development in the frontend (including mobile) environment by
a team of the fixed supplier for frontend development (offsite, Belgium). Project manager of the
Telecom company onsite.
: 1
 BSCS CDR Cycling Control. Small enhancement on a Billing application. Performed by a projecttea
m (globally distributed; partly offsite, Indian supplier) for Billing system enhancements. Perfo
rmed as a  fixed price project.
: 1
 Campaign Management - Change Request 2. Sub-project of a large implementation of a new Campaign
Management system in the CX department (Customer Experience). Plan-driven development approach.
Globally distributed team: Project Manager, Business Analyst, Main Developer, Testers onsite, de
velopers offsite (India). High management attention due to regular budget and schedule overrun.
Project suffered from a bad imago in Telecom organisation due to ongoing problems and many defec
ts. : 1
 Campaign Management Tool (Comviva) WP05. Sub-project of a large implementation of a new Campaig
n Management system in the CX department (Customer Experience). Plan-driven development approach
. Globally distributed team: Project Manager, Business Analyst, Main Developer, Testers onsite,
developers offsite (India). High management attention due to regular budget and schedule overrun
. Project suffered from a bad image in Telecom organisation due to ongoing problems and many def
ects.: 1
 (Other)
:16
      Size            Duration          Cost          CostPerFP        DaysPerFP
 Min.   : 12.00   Min.   : 4.96   Min.   :  8000   Min.   : 434.0   Min.   : 0.870
 1st Qu.: 24.25   1st Qu.: 8.95   1st Qu.: 46013   1st Qu.: 836.8   1st Qu.: 2.635
 Median : 38.50   Median :10.29   Median : 66209   Median :1612.0   Median : 7.850
 Mean   : 80.77   Mean   :10.22   Mean   : 99562   Mean   :1932.2   Mean   : 9.647
 3rd Qu.:127.25   3rd Qu.:11.50   3rd Qu.:135342   3rd Qu.:2790.2   3rd Qu.:15.440
 Max.   :324.00   Max.   :19.03   Max.   :296000   Max.   :4976.0   Max.   :27.840

  DefectsPerFP    CostDurationIndex           Quadrant     EQFCost
 Min.   :0.0200   Min.   :86.92     Bad Practice  : 6   Min.   :0.000
 1st Qu.:0.0975   1st Qu.:93.27     Cost over Time:14   1st Qu.:2.123
 Median :0.2150   Median :97.28     Good Practice : 2   Median :3.520
 Mean   :0.3728   Mean   :95.90                         Mean   :4.321
 3rd Qu.:0.6000   3rd Qu.:98.57                         3rd Qu.:7.992
 Max.   :1.4400   Max.   :99.78                         Max.   :9.140
 NA's   :4                                              NA's   :10
  EQFDuration     MultiApplicationRelease  ReleaseBased    OnceOnlyProject
 Min.   : 1.890   Min.   :0.0000          Min.   :0.0000   Min.   :0.0000
 1st Qu.: 2.393   1st Qu.:0.0000          1st Qu.:0.0000   1st Qu.:0.0000
 Median : 3.600   Median :0.0000          Median :0.0000   Median :0.0000
 Mean   : 7.273   Mean   :0.2727          Mean   :0.2727   Mean   :0.4545
 3rd Qu.: 6.327   3rd Qu.:0.7500          3rd Qu.:0.7500   3rd Qu.:1.0000
 Max.   :27.570   Max.   :1.0000          Max.   :1.0000   Max.   :1.0000
 NA's   :2
 Single.application Dependencies.with.other.applications Migration.Project
 Min.   :0.0000     Min.   :0.0000                        Min.   :0.00000
 1st Qu.:0.0000     1st Qu.:0.0000                        1st Qu.:0.00000
 Median :0.0000     Median :0.0000                        Median :0.00000
 Mean   :0.4545     Mean   :0.4091                        Mean   :0.04545
 3rd Qu.:1.0000     3rd Qu.:1.0000                        3rd Qu.:0.00000
 Max.   :1.0000     Max.   :1.0000                        Max.   :1.00000
```

```
 PackageOffTheShelf SteadyHeartbeat  FixedExperiencedTeam BusinessDriven
 Min.   :0          Min.   :0.0000   Min.   :0.0000       Min.   :0.0000
 1st Qu.:0          1st Qu.:0.0000   1st Qu.:0.0000       1st Qu.:1.0000
 Median :0          Median :0.0000   Median :0.0000       Median :1.0000
 Mean   :0          Mean   :0.2273   Mean   :0.2273       Mean   :0.8182
 3rd Qu.:0          3rd Qu.:0.0000   3rd Qu.:0.0000       3rd Qu.:1.0000
 Max.   :0          Max.   :1.0000   Max.   :1.0000       Max.   :1.0000

 RulesAndRegulationsDriven   PilotOrPOC   YearGoLive
 Min.   :0.0000            Min.   :0    Min.   :2014
 1st Qu.:0.0000            1st Qu.:0    1st Qu.:2015
 Median :0.0000            Median :0    Median :2015
 Mean   :0.1364            Mean   :0    Mean   :2015
 3rd Qu.:0.0000            3rd Qu.:0    3rd Qu.:2015
 Max.   :1.0000            Max.   :0    Max.   :2015

                     BusinessDomain   DevelopmentMethod
 Billing                       :7      Plan-driven:16
 Client and Account Management:8      Scrum      : 6
 Data Warehouse & BI           :2
 Internet & Mobile             :5



 StakeholderSatisfactionProcess Stakeholder.SatisfactionProduct
 Min.   :2.500                  Min.   :2.500
 1st Qu.:3.000                  1st Qu.:3.200
 Median :3.500                  Median :3.630
 Mean   :3.521                  Mean   :3.605
 3rd Qu.:4.000                  3rd Qu.:4.000
 Max.   :4.500                  Max.   :4.500
 NA's   :1                      NA's   :1
 PerceivedValueCustomer PerceivedValueFinancial PerceivedValueInternalProces
 Min.   :2.610          Min.   :2.610           Min.   :2.610
 1st Qu.:3.203          1st Qu.:3.203           1st Qu.:3.203
 Median :3.330          Median :3.330           Median :3.330
 Mean   :3.357          Mean   :3.357           Mean   :3.357
 3rd Qu.:3.645          3rd Qu.:3.645           3rd Qu.:3.645
 Max.   :4.000          Max.   :4.000           Max.   :4.000
 NA's   :2              NA's   :2               NA's   :2
 PerceivedValueInnovation PerceivedValueOverall    Defects       ErrorsFirstMonth
 Min.   :2.610            Min.   :2.610         Min.   :  1.00   Min.   :1
 1st Qu.:3.203            1st Qu.:3.203         1st Qu.:  3.25   1st Qu.:1
 Median :3.330            Median :3.330         Median :  8.50   Median :1
 Mean   :3.357            Mean   :3.357         Mean   : 33.61   Mean   :1
 3rd Qu.:3.645            3rd Qu.:3.645         3rd Qu.: 26.50   3rd Qu.:1
 Max.   :4.000            Max.   :4.000         Max.   :223.00   Max.   :1
 NA's   :2                NA's   :2             NA's   :4        NA's   :21
                 DevelopmentClassification
 Conversion (<5% new)          : 3
 Major Enhancement (25-75% new): 7
 Minor Enhancement (5-25% new) : 1
 New Development               :11
```

7.2.2 Summary of the EBSPM repository

```
> summary(EBSPM)
       ID           Organization ProjectID
 Min.   :  1.0   Banking A:206   Mode:logical
 1st Qu.:123.8   Banking B:125   NA's:492
```

```
Median :246.5   Beltel   :157
Mean   :246.5   DutchCo  :  4
3rd Qu.:369.2
Max.   :492.0

                                  ProjectDescription      Size
Customer Lifetime Value                   :  2      Min.   :   4.00
Remove duplicate activation in bundle area:  2      1st Qu.:  38.75
Replacement of the roaming tariff solution:  2      Median : 116.00
&V: Renewal Search Engine (GICT)          :  1      Mean   : 220.26
(PCK) Plan Releases (GICT)                :  1      3rd Qu.: 250.25
1-day porting (legal/BIPT requirement)    :  1      Max.   :4600.00
(Other)                                   :483
   Duration           Cost           CostPerFP       DaysPerFP
Min.   : 0.900   Min.   :    329   Min.   :   82   Min.   : 0.1100
1st Qu.: 5.397   1st Qu.:  70796   1st Qu.: 1236   1st Qu.: 0.9975
Median : 8.411   Median : 278156   Median : 2520   Median : 2.0800
Mean   : 8.950   Mean   : 592802   Mean   : 3618   Mean   : 5.5049
3rd Qu.:11.070   3rd Qu.: 699527   3rd Qu.: 4546   3rd Qu.: 5.4275
Max.   :26.840   Max.   :6802466   Max.   :35818   Max.   :67.2500

 DefectsPerFP    CostDurationIndex           Quadrant   EQFCost
Min.   :0.0000   Min.   : 6.19     Bad Practice  :167   Mode:logical
1st Qu.:0.0800   1st Qu.:86.69     Cost over Time:104   NA's:492
Median :0.1600   Median :91.72     Good Practice :144
Mean   :0.3513   Mean   :89.19     Time over Cost: 77
3rd Qu.:0.3600   3rd Qu.:95.46
Max.   :5.3400   Max.   :99.14
NA's   :203
EQFDuration    MultiApplicationRelease ReleaseBased   OnceOnlyProject
Mode:logical   Mode:logical            Mode:logical   Mode:logical
NA's:492       NA's:492                NA's:492       NA's:492




Single.application Dependencies.with.other.applications Migration.Project
Mode:logical       Mode:logical                          Mode:logical
NA's:492           NA's:492                              NA's:492




PackageOffTheShelf SteadyHeartbeat FixedExperiencedTeam BusinessDriven
Mode:logical       Mode:logical    Mode:logical         Mode:logical
NA's:492           NA's:492        NA's:492             NA's:492




RulesAndRegulationsDriven PilotOrPOC       YearGoLive
Mode:logical              Mode:logical   Min.   :2008
NA's:492                  NA's:492       1st Qu.:2010
                                         Median :2011
                                         Mean   :2012
                                         3rd Qu.:2014
                                         Max.   :2016
```

```
                      BusinessDomain   DevelopmentMethod
 Client and Account Management:107    Plan-driven:385
 Internet & Mobile            : 71    RUP        :  1
 Finance & Risk               : 54    Scrum      :106
 Payments                     : 50
 Billing                      : 46
 Savings & Loans              : 40
 (Other)                      :124
 StakeholderSatisfactionProcess Stakeholder.SatisfactionProduct
 Mode:logical                    Mode:logical
 NA's:492                        NA's:492




 PerceivedValueCustomer PerceivedValueFinancial PerceivedValueInternalProces
 Mode:logical           Mode:logical            Mode:logical
 NA's:492               NA's:492                NA's:492




 PerceivedValueInnovation PerceivedValueOverall    Defects        ErrorsFirstMonth
 Mode:logical             Mode:logical          Min.   :   0.00   Min.   :  0.000
 NA's:492                 NA's:492              1st Qu.:   7.00   1st Qu.:  0.000
                                                Median :  21.00   Median :  1.000
                                                Mean   :  71.74   Mean   :  8.833
                                                3rd Qu.:  58.25   3rd Qu.:  3.000
                                                Max.   :1586.00   Max.   :200.000
                                                NA's   :222       NA's   :432
                 DevelopmentClassification
 Conversion (<5% new)          : 28
 Major Enhancement (25-75% new):156
 Minor Enhancement (5-25% new) :126
 New Development               :182
```

7.2.3 Summary of the DutchCo subset

```
> summary(RepositoryDutchco)
       ID        Organization           ProjectID
 Min.   :446.0   Beltel:0     DutchCo 5.1 - POR:1
 1st Qu.:446.8   DutchCo:4    DutchCo 5.1 -AM  :1
 Median :469.0                DutchCo 5.2      :1
 Mean   :469.0                DutchCo 5.3      :1
 3rd Qu.:491.2                Project 3.1    :0
 Max.   :492.0                Project 3.2    :0
                              (Other)        :0

ProjectDescription
 DutchCo Release 5.2
:3
 DutchCo Release 5.3 (Baykom)
:1
 Adapt T&C Procedure DOB Wmode Platform. Small enhancement to an existing system. Plan-driven de
livery approach. Performed by a team of the strategic partner (Indian, partly onsite, partly off
site). Project manager of the Telecom company (onsite).:0
```

```
 Admin Fee for SoHo Customers. Rules- and regulations driven small enhancement to existing CRM s
ystem. Small project team of onsite Project Manager and Business Analyst and offsite Indian deve
lopers.                                                 :0
 Apple Store SSO and QR Code. New development in the frontend (including mobile) environment by
a team of the fixed supplier for frontend development (offsite, Belgium). Project manager of the
Telecom company onsite.                                 :0
 BSCS CDR Cycling Control. Small enhancement on a Billing application. Performed by a projecttea
m (globally distributed; partly offsite, Indian supplier) for Billing system enhancements. Perfo
rmed as a  fixed price project.                         :0
 (Other)
:0
      Size           Duration           Cost          CostPerFP       DaysPerFP
 Min.   : 277.0   Min.   : 3.780   Min.   :125827   Min.   :376.0   Min.   :0.1600
 1st Qu.: 320.5   1st Qu.: 5.902   1st Qu.:192040   1st Qu.:589.8   1st Qu.:0.1975
 Median : 784.0   Median : 7.710   Median :514486   Median :686.0   Median :0.2750
 Mean   : 776.5   Mean   : 7.372   Mean   :512897   Mean   :630.2   Mean   :0.4600
 3rd Qu.:1240.0   3rd Qu.: 9.180   3rd Qu.:835343   3rd Qu.:726.5   3rd Qu.:0.5375
 Max.   :1261.0   Max.   :10.290   Max.   :896788   Max.   :773.0   Max.   :1.1300

  DefectsPerFP    CostDurationIndex           Quadrant    EQFCost
 Min.   :0.0400   Min.   :95.03     Bad Practice  :0   Min.   : NA
 1st Qu.:0.1075   1st Qu.:96.12     Cost over Time:0   1st Qu.: NA
 Median :0.2200   Median :96.81     Good Practice :4   Median : NA
 Mean   :0.2100   Mean   :96.72                        Mean   :NaN
 3rd Qu.:0.3225   3rd Qu.:97.41                        3rd Qu.: NA
 Max.   :0.3600   Max.   :98.22                        Max.   : NA
                                                       NA's   :4
  EQFDuration     MultiApplicationRelease  ReleaseBased OnceOnlyProject
 Min.   :0.8100   Min.   :0               Min.   :1     Min.   :0
 1st Qu.:0.9075   1st Qu.:0               1st Qu.:1     1st Qu.:0
 Median :1.0050   Median :0               Median :1     Median :0
 Mean   :1.0050   Mean   :0               Mean   :1     Mean   :0
 3rd Qu.:1.1025   3rd Qu.:0               3rd Qu.:1     3rd Qu.:0
 Max.   :1.2000   Max.   :0               Max.   :1     Max.   :0
 NA's   :2
 Single.application Dependencies.with.other.applications Migration.Project
 Min.   :1          Min.   :0                             Min.   :0
 1st Qu.:1          1st Qu.:0                             1st Qu.:0
 Median :1          Median :0                             Median :0
 Mean   :1          Mean   :0                             Mean   :0
 3rd Qu.:1          3rd Qu.:0                             3rd Qu.:0
 Max.   :1          Max.   :0                             Max.   :0

 PackageOffTheShelf SteadyHeartbeat FixedExperiencedTeam BusinessDriven
 Min.   :0          Min.   :1       Min.   :1            Min.   :1
 1st Qu.:0          1st Qu.:1       1st Qu.:1            1st Qu.:1
 Median :0          Median :1       Median :1            Median :1
 Mean   :0          Mean   :1       Mean   :1            Mean   :1
 3rd Qu.:0          3rd Qu.:1       3rd Qu.:1            3rd Qu.:1
 Max.   :0          Max.   :1       Max.   :1            Max.   :1

 RulesAndRegulationsDriven   PilotOrPOC   YearGoLive
 Min.   :0                 Min.   :0    Min.   :2016
 1st Qu.:0                 1st Qu.:0    1st Qu.:2016
 Median :0                 Median :0    Median :2016
 Mean   :0                 Mean   :0    Mean   :2016
 3rd Qu.:0                 3rd Qu.:0    3rd Qu.:2016
 Max.   :0                 Max.   :0    Max.   :2016

                     BusinessDomain   DevelopmentMethod
```

```
 Billing                       :4      Plan-driven:0
 Client and Account Management:0      Scrum      :4
 Data Warehouse & BI           :0
 Internet & Mobile             :0



 StakeholderSatisfactionProcess Stakeholder.SatisfactionProduct
 Min.   :3.33                   Min.   :3.88
 1st Qu.:3.33                   1st Qu.:3.88
 Median :3.33                   Median :3.88
 Mean   :3.33                   Mean   :3.88
 3rd Qu.:3.33                   3rd Qu.:3.88
 Max.   :3.33                   Max.   :3.88
 NA's   :3                      NA's   :3
 PerceivedValueCustomer PerceivedValueFinancial PerceivedValueInternalProces
 Min.   :4.04           Min.   :3.89            Min.   :3.52
 1st Qu.:4.04           1st Qu.:3.89            1st Qu.:3.52
 Median :4.04           Median :3.89            Median :3.52
 Mean   :4.04           Mean   :3.89            Mean   :3.52
 3rd Qu.:4.04           3rd Qu.:3.89            3rd Qu.:3.52
 Max.   :4.04           Max.   :3.89            Max.   :3.52
 NA's   :3              NA's   :3               NA's   :3
 PerceivedValueInnovation PerceivedValueOverall    Defects      ErrorsFirstMonth
 Min.   :4.17               Min.   :3.91         Min.   : 15.0   Min.   : NA
 1st Qu.:4.17               1st Qu.:3.91         1st Qu.: 78.0   1st Qu.: NA
 Median :4.17               Median :3.91         Median :131.0   Median : NA
 Mean   :4.17               Mean   :3.91         Mean   :165.8   Mean   :NaN
 3rd Qu.:4.17               3rd Qu.:3.91         3rd Qu.:218.8   3rd Qu.: NA
 Max.   :4.17               Max.   :3.91         Max.   :386.0   Max.   : NA
 NA's   :3                  NA's   :3                            NA's   :4
                 DevelopmentClassification
 Conversion (<5% new)          :0
 Major Enhancement (25-75% new):4
 Minor Enhancement (5-25% new) :0
 New Development               :0
```

7.2.4 Descriptive statistics of the DutchCo subset

```
> str(RepositoryDutchco)
'data.frame':    4 obs. of  39 variables:
 $ ID                                : int  446 447 491 492
 $ Organization                      : Factor w/ 2 levels "Beltel"," DutchCo ": 2 2 2 2
 $ ProjectID                         : Factor w/ 26 levels " DutchCo 5.1 - POR",..: 2 1 3 4
 $ ProjectDescription                : Factor w/ 24 levels "Adapt T&C Procedure DOB Wmode Plat
form. Small enhancement to an existing system. Plan-driven delivery approach. Performed by a "|
__truncated__,..: 5 5 5 6
 $ Size                              : int  335 277 1233 1261
 $ Duration                          : num  3.78 10.29 6.61 8.81
 $ Cost                              : int  125827 214111 814861 896788
 $ CostPerFP                         : int  376 773 661 711
 $ DaysPerFP                         : num  0.34 1.13 0.16 0.21
 $ DefectsPerFP                      : num  0.04 0.36 0.31 0.13
 $ CostDurationIndex                 : num  98.2 95 97.1 96.5
 $ Quadrant                          : Factor w/ 3 levels "Bad Practice",..: 3 3 3 3
 $ EQFCost                           : num  NA NA NA NA
 $ EQFDuration                       : num  NA NA 0.81 1.2
 $ MultiApplicationRelease           : int  0 0 0 0
 $ ReleaseBased                      : int  1 1 1 1
```

```
 $ OnceOnlyProject                     : int  0 0 0 0
 $ Single.application                  : int  1 1 1 1
 $ Dependencies.with.other.applications: int  0 0 0 0
 $ Migration.Project                   : int  0 0 0 0
 $ PackageOffTheShelf                  : int  0 0 0 0
 $ SteadyHeartbeat                     : int  1 1 1 1
 $ FixedExperiencedTeam                : int  1 1 1 1
 $ BusinessDriven                      : int  1 1 1 1
 $ RulesAndRegulationsDriven           : int  0 0 0 0
 $ PilotOrPOC                          : int  0 0 0 0
 $ YearGoLive                          : int  2016 2016 2016 2016
 $ BusinessDomain                      : Factor w/ 4 levels "Billing","Client and Account Manage
ment",..: 1 1 1 1
 $ DevelopmentMethod                   : Factor w/ 2 levels "Plan-driven",..: 2 2 2 2
 $ StakeholderSatisfactionProcess      : num  NA NA NA 3.33
 $ Stakeholder.SatisfactionProduct     : num  NA NA NA 3.88
 $ PerceivedValueCustomer              : num  NA NA NA 4.04
 $ PerceivedValueFinancial             : num  NA NA NA 3.89
 $ PerceivedValueInternalProces        : num  NA NA NA 3.52
 $ PerceivedValueInnovation            : num  NA NA NA 4.17
 $ PerceivedValueOverall               : num  NA NA NA 3.91
 $ Defects                             : int  15 99 386 163
 $ ErrorsFirstMonth                    : int  NA NA NA NA
 $ DevelopmentClassification           : Factor w/ 4 levels "Conversion (<5% new)",..: 2 2 2 2
```

7.2.5 Descriptive statistics of the BelTel subset

```
> str(RepositoryBeltel)
'data.frame':    22 obs. of  39 variables:
 $ ID                                  : int  390 393 403 411 423 424 425 426 427 428 ...
 $ Organization                        : Factor w/ 2 levels "Beltel"," DutchCo ": 1 1 1 1 1 1 1
1 1 1 ...
 $ ProjectID                           : Factor w/ 26 levels " DutchCo 5.1 - POR",..: 9 6 8 10 7
5 13 11 12 14 ...
 $ ProjectDescription                  : Factor w/ 24 levels "Adapt T&C Procedure DOB Wmode Plat
form. Small enhancement to an existing system. Plan-driven delivery approach. Performed by a "|
__truncated__,..: 20 4 7 24 9 14 2 10 13 11 ...
 $ Size                                : int  232 34 177 35 128 13 12 63 130 28 ...
 $ Duration                            : num  12.59 9.69 9.69 19.03 8.81 ...
 $ Cost                                : int  100752 25092 143575 110643 62964 45817 54500 11000
0 210300 37455 ...
 $ CostPerFP                           : int  434 738 811 3161 492 3524 4542 1746 1618 1338 ...
 $ DaysPerFP                           : num  1.65 8.67 1.67 16.55 2.09 ...
 $ DefectsPerFP                        : num  0.22 0.09 0.8 0.86 0.13 0.92 0.33 0.16 0.02 0.21 .
..
 $ CostDurationIndex                   : num  99.8 97.5 86.9 98.8 97.3 ...
 $ Quadrant                            : Factor w/ 3 levels "Bad Practice",..: 2 2 2 1 2 1 1 3 2
2 ...
 $ EQFCost                             : num  NA NA NA 0 NA NA 3.94 0 NA 0 ...
 $ EQFDuration                         : num  1.89 2.68 3.99 3.11 NA ...
 $ MultiApplicationRelease             : int  1 1 1 1 0 0 1 0 0 0 ...
 $ ReleaseBased                        : int  0 0 0 0 1 1 0 1 0 1 ...
 $ OnceOnlyProject                     : int  0 0 0 0 0 0 0 0 1 0 ...
 $ Single.application                  : int  1 0 0 0 0 0 0 1 0 1 ...
 $ Dependencies.with.other.applications: int  0 1 1 1 1 1 1 0 1 0 ...
 $ Migration.Project                   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ PackageOffTheShelf                  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ SteadyHeartbeat                     : int  0 0 0 0 0 0 0 1 0 1 ...
 $ FixedExperiencedTeam                : int  0 0 0 0 0 0 0 1 0 1 ...
```

```
 $ BusinessDriven                      : int  1 1 1 1 0 0 1 1 1 1 ...
 $ RulesAndRegulationsDriven           : int  0 0 0 0 1 1 0 0 0 0 ...
 $ PilotOrPOC                          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ YearGoLive                          : int  2014 2015 2015 2015 2015 2015 2015 2015 2015 2015
...
 $ BusinessDomain                      : Factor w/ 4 levels "Billing","Client and Account Manage
ment",..: 4 1 3 2 2 2 1 2 4 2 ...
 $ DevelopmentMethod                   : Factor w/ 2 levels "Plan-driven",..: 1 1 2 1 2 1 1 2 1
2 ...
 $ StakeholderSatisfactionProcess      : num  3.7 3.63 2.83 2.5 NA 3.5 3.4 3.92 4.19 4.5 ...
 $ Stakeholder.SatisfactionProduct     : num  3.7 3.63 2.83 2.5 NA 3.5 3.4 3.92 4.19 4.5 ...
 $ PerceivedValueCustomer              : num  3.83 3.44 3.63 3.06 NA 3.58 2.61 3.3 3.86 4 ...
 $ PerceivedValueFinancial             : num  3.83 3.44 3.63 3.06 NA 3.58 2.61 3.3 3.86 4 ...
 $ PerceivedValueInternalProces        : num  3.83 3.44 3.63 3.06 NA 3.58 2.61 3.3 3.86 4 ...
 $ PerceivedValueInnovation            : num  3.83 3.44 3.63 3.06 NA 3.58 2.61 3.3 3.86 4 ...
 $ PerceivedValueOverall               : num  3.83 3.44 3.63 3.06 NA 3.58 2.61 3.3 3.86 4 ...
 $ Defects                             : int  51 3 142 30 16 12 4 10 3 6 ...
 $ ErrorsFirstMonth                    : int  NA NA NA 1 NA NA NA NA NA NA ...
 $ DevelopmentClassification           : Factor w/ 4 levels "Conversion (<5% new)",..: 2 4 1 1 4
4 4 4 4 4 ...
```

7.2.6 Descriptive statistics of all BelTel projects in the EBSPM repository

```
> str(RepositoryBeltelAll)
'data.frame':    157 obs. of  39 variables:
 $ ID                                  : int  312 313 314 315 316 317 318 319 320 321 ...
 $ Organization                        : Factor w/ 4 levels "Banking A","Banking B",..: 3 3 3 3
3 3 3 3 3 3 ...
 $ ProjectID                           : logi  NA NA NA NA NA NA ...
 $ ProjectDescription                  : Factor w/ 489 levels "&V: Renewal Search Engine (GICT)"
,..: 95 66 123 471 262 122 80 460 302 116 ...
 $ Size                                : num  17 24 1089 341 238 ...
 $ Duration                            : num  13.74 15.42 24.16 14.45 4.13 ...
 $ Cost                                : num  113040 382790 1600000 241400 140000 ...
 $ CostPerFP                           : int  6649 15950 1469 708 588 563 3127 5431 3195 2600 ..
.
 $ DaysPerFP                           : num  24.6 19.55 0.68 1.29 0.53 ...
 $ DefectsPerFP                        : num  NA NA 0.13 NA NA 0.07 0.24 NA NA NA ...
 $ CostDurationIndex                   : num  78.1 57.9 87.8 92.6 97.8 ...
 $ Quadrant                            : Factor w/ 4 levels "Bad Practice",..: 1 1 2 2 3 2 4 1 1
2 ...
 $ EQFCost                             : logi  NA NA NA NA NA NA ...
 $ EQFDuration                         : logi  NA NA NA NA NA NA ...
 $ MultiApplicationRelease             : logi  NA NA NA NA NA NA ...
 $ ReleaseBased                        : logi  NA NA NA NA NA NA ...
 $ OnceOnlyProject                     : logi  NA NA NA NA NA NA ...
 $ Single.application                  : logi  NA NA NA NA NA NA ...
 $ Dependencies.with.other.applications: logi  NA NA NA NA NA NA ...
 $ Migration.Project                   : logi  NA NA NA NA NA NA ...
 $ PackageOffTheShelf                  : logi  NA NA NA NA NA NA ...
 $ SteadyHeartbeat                     : logi  NA NA NA NA NA NA ...
 $ FixedExperiencedTeam                : logi  NA NA NA NA NA NA ...
 $ BusinessDriven                      : logi  NA NA NA NA NA NA ...
 $ RulesAndRegulationsDriven           : logi  NA NA NA NA NA NA ...
 $ PilotOrPOC                          : logi  NA NA NA NA NA NA ...
 $ YearGoLive                          : int  2011 2011 2012 2013 2013 2012 2012 2011 2013 2012
...
 $ BusinessDomain                      : Factor w/ 11 levels "Billing","Call Center Solutions",.
.: 3 3 3 7 7 4 3 3 3 3 ...
```

```
 $ DevelopmentMethod                   : Factor w/ 3 levels "Plan-driven",..: 1 1 1 1 1 1 1 1 1
1 ...
 $ StakeholderSatisfactionProcess      : logi  NA NA NA NA NA NA ...
 $ Stakeholder.SatisfactionProduct     : logi  NA NA NA NA NA NA ...
 $ PerceivedValueCustomer              : logi  NA NA NA NA NA NA ...
 $ PerceivedValueFinancial             : logi  NA NA NA NA NA NA ...
 $ PerceivedValueInternalProces        : logi  NA NA NA NA NA NA ...
 $ PerceivedValueInnovation            : logi  NA NA NA NA NA NA ...
 $ PerceivedValueOverall               : logi  NA NA NA NA NA NA ...
 $ Defects                             : int  NA NA 131 NA NA 35 107 NA NA NA ...
 $ ErrorsFirstMonth                    : int  NA NA 14 NA NA NA NA NA NA NA ...
 $ DevelopmentClassification           : Factor w/ 4 levels "Conversion (<5% new)",..: 2 4 4 3 4
2 4 3 3 2 ...
```

7.2.7 Descriptive statistics of all DutchCo projects in the EBSPM repository

```
> str(RepositoryDutchcoAll)
'data.frame':    4 obs. of  39 variables:
 $ ID                                  : int  446 447 491 492
 $ Organization                        : Factor w/ 4 levels "Banking A","Banking B",..: 4 4 4 4
 $ ProjectID                           : logi  NA NA NA NA
 $ ProjectDescription                  : Factor w/ 489 levels "&V: Renewal Search Engine (GICT)"
,..: 73 74 75 76
 $ Size                                : num  335 277 1233 1261
 $ Duration                            : num  3.78 10.29 6.61 8.81
 $ Cost                                : num  125827 214111 814861 896788
 $ CostPerFP                           : int  376 773 661 711
 $ DaysPerFP                           : num  0.34 1.13 0.16 0.21
 $ DefectsPerFP                        : num  0.04 0.36 0.31 0.13
 $ CostDurationIndex                   : num  98.2 95 97.1 96.5
 $ Quadrant                            : Factor w/ 4 levels "Bad Practice",..: 3 2 3 3
 $ EQFCost                             : logi  NA NA NA NA
 $ EQFDuration                         : logi  NA NA NA NA
 $ MultiApplicationRelease             : logi  NA NA NA NA
 $ ReleaseBased                        : logi  NA NA NA NA
 $ OnceOnlyProject                     : logi  NA NA NA NA
 $ Single.application                  : logi  NA NA NA NA
 $ Dependencies.with.other.applications: logi  NA NA NA NA
 $ Migration.Project                   : logi  NA NA NA NA
 $ PackageOffTheShelf                  : logi  NA NA NA NA
 $ SteadyHeartbeat                     : logi  NA NA NA NA
 $ FixedExperiencedTeam                : logi  NA NA NA NA
 $ BusinessDriven                      : logi  NA NA NA NA
 $ RulesAndRegulationsDriven           : logi  NA NA NA NA
 $ PilotOrPOC                          : logi  NA NA NA NA
 $ YearGoLive                          : int  2015 2015 2015 2015
 $ BusinessDomain                      : Factor w/ 11 levels "Billing","Call Center Solutions",.
.: 4 4 4 4
 $ DevelopmentMethod                   : Factor w/ 3 levels "Plan-driven",..: 3 3 3 3
 $ StakeholderSatisfactionProcess      : logi  NA NA NA NA
 $ Stakeholder.SatisfactionProduct     : logi  NA NA NA NA
 $ PerceivedValueCustomer              : logi  NA NA NA NA
 $ PerceivedValueFinancial             : logi  NA NA NA NA
 $ PerceivedValueInternalProces        : logi  NA NA NA NA
 $ PerceivedValueInnovation            : logi  NA NA NA NA
 $ PerceivedValueOverall               : logi  NA NA NA NA
 $ Defects                             : int  15 99 386 163
 $ ErrorsFirstMonth                    : int  NA NA NA NA
 $ DevelopmentClassification           : Factor w/ 4 levels "Conversion (<5% new)",..: 2 2 2 2
```

7.2.8 Descriptive statistics of the EBSPM repository

```
> str(EBSPM)
'data.frame':    492 obs. of  39 variables:
 $ ID                                : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Organization                      : Factor w/ 4 levels "Banking A","Banking B",..: 2 2 2 2
2 2 2 2 2 2 ...
 $ ProjectID                         : logi  NA NA NA NA NA NA ...
 $ ProjectDescription                : Factor w/ 489 levels "&V: Renewal Search Engine (GICT)"
,..: 316 363 348 210 357 489 382 356 103 181 ...
 $ Size                              : num  47 75 128 131 93 15 107 56 48 469 ...
 $ Duration                          : num  4.9 3.81 7.48 4.38 4.19 ...
 $ Cost                              : num  489620 343250 989570 1115640 187880 ...
 $ CostPerFP                         : int  10417 4577 7731 8516 2020 24924 1983 8185 6862 238
7 ...
 $ DaysPerFP                         : num  3.17 1.55 1.78 1.02 1.37 ...
 $ DefectsPerFP                      : num  NA NA NA NA 0.02 NA NA NA NA NA ...
 $ CostDurationIndex                 : num  77.9 90.8 80.7 79.9 95.8 ...
 $ Quadrant                          : Factor w/ 4 levels "Bad Practice",..: 4 4 4 4 3 1 3 4 4
1 ...
 $ EQFCost                           : logi  NA NA NA NA NA NA ...
 $ EQFDuration                       : logi  NA NA NA NA NA NA ...
 $ MultiApplicationRelease           : logi  NA NA NA NA NA NA ...
 $ ReleaseBased                      : logi  NA NA NA NA NA NA ...
 $ OnceOnlyProject                   : logi  NA NA NA NA NA NA ...
 $ Single.application                : logi  NA NA NA NA NA NA ...
 $ Dependencies.with.other.applications: logi  NA NA NA NA NA NA ...
 $ Migration.Project                 : logi  NA NA NA NA NA NA ...
 $ PackageOffTheShelf                : logi  NA NA NA NA NA NA ...
 $ SteadyHeartbeat                   : logi  NA NA NA NA NA NA ...
 $ FixedExperiencedTeam              : logi  NA NA NA NA NA NA ...
 $ BusinessDriven                    : logi  NA NA NA NA NA NA ...
 $ RulesAndRegulationsDriven         : logi  NA NA NA NA NA NA ...
 $ PilotOrPOC                        : logi  NA NA NA NA NA NA ...
 $ YearGoLive                        : int  2009 2008 2009 2008 2008 2009 2009 2009 2009 2009
...
 $ BusinessDomain                    : Factor w/ 11 levels "Billing","Call Center Solutions",.
.: 10 10 10 11 11 11 11 3 10 5 ...
 $ DevelopmentMethod                 : Factor w/ 3 levels "Plan-driven",..: 1 1 1 1 1 1 1 1 1
1 ...
 $ StakeholderSatisfactionProcess    : logi  NA NA NA NA NA NA ...
 $ Stakeholder.SatisfactionProduct   : logi  NA NA NA NA NA NA ...
 $ PerceivedValueCustomer            : logi  NA NA NA NA NA NA ...
 $ PerceivedValueFinancial           : logi  NA NA NA NA NA NA ...
 $ PerceivedValueInternalProces      : logi  NA NA NA NA NA NA ...
 $ PerceivedValueInnovation          : logi  NA NA NA NA NA NA ...
 $ PerceivedValueOverall             : logi  NA NA NA NA NA NA ...
 $ Defects                           : int  NA NA NA NA 2 NA NA NA NA NA ...
 $ ErrorsFirstMonth                  : int  NA NA NA NA NA NA NA NA NA NA ...
 $ DevelopmentClassification         : Factor w/ 4 levels "Conversion (<5% new)",..: 4 4 4 2 4
3 3 2 3 4 ...
```

7.2.9 Example of the outcome of correlation testing between BelTel Cost and BelTel Size

```
> cor.test(RepositoryBeltel$Cost, RepositoryBeltel$Size,
+          use='pairwise', alternative = c("two.sided", "less", "greater"),
+          method = c("spearman"),
+          exact = NULL, conf.level = 0.95, continuity = FALSE)
```

```
        Spearman's rank correlation rho

data:  RepositoryBeltel$Cost and RepositoryBeltel$Size
S = 467.53, p-value = 9.442e-05
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.7360092

Warning message:
In cor.test.default(RepositoryBeltel$Cost, RepositoryBeltel$Size,  :
  Cannot compute exact p-value with ties
```

7.2.10 Example of the outcome of Wilcoxon test between EBSPM Size and Repository Size

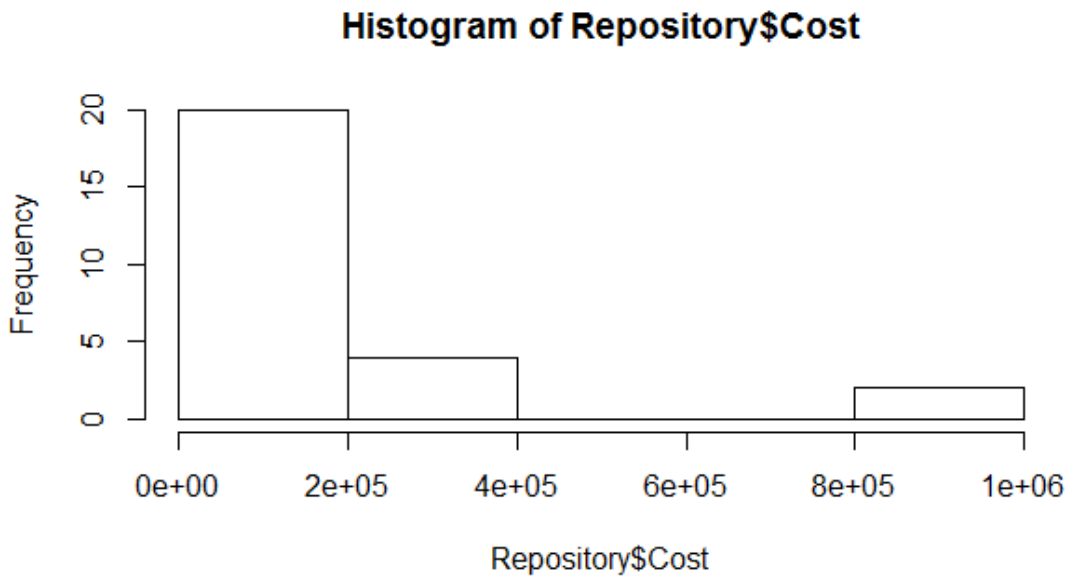
```
> wilcox.test(EBSPM$Size,Repository$Size)

        Wilcoxon rank sum test with continuity correction

data:  EBSPM$Size and Repository$Size
W = 7367.5, p-value = 0.1917
alternative hypothesis: true location shift is not equal to 0
```

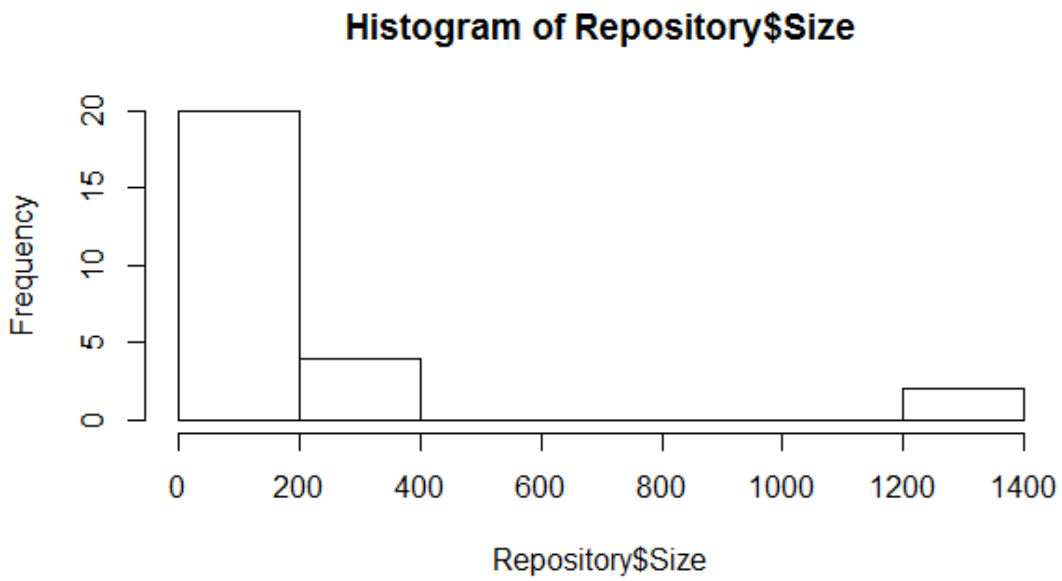
7.2.11 Outcome of hist(Repository$Cost)

```
> hist(Repository$Cost)
```



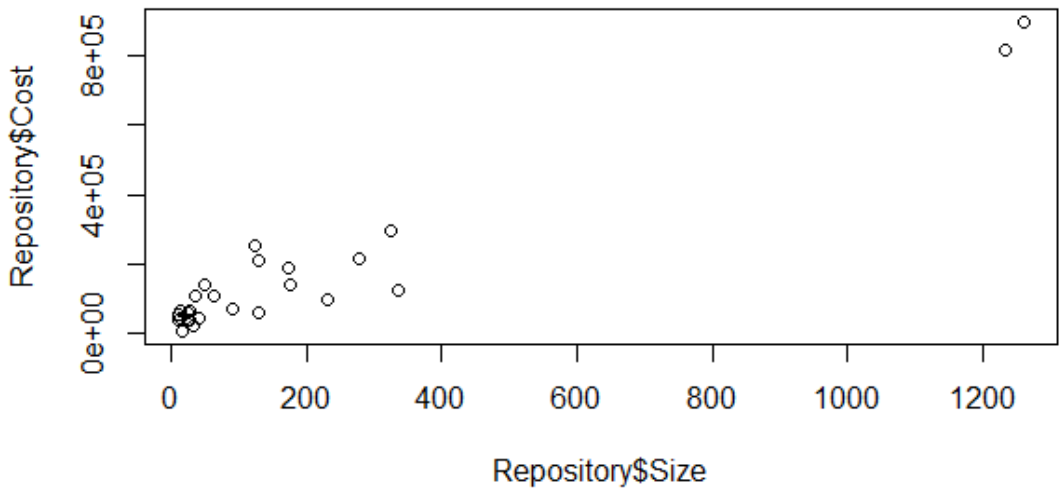
7.2.12 Outcome of hist(Repository$Size)

```
> hist(Repository$Size)
```



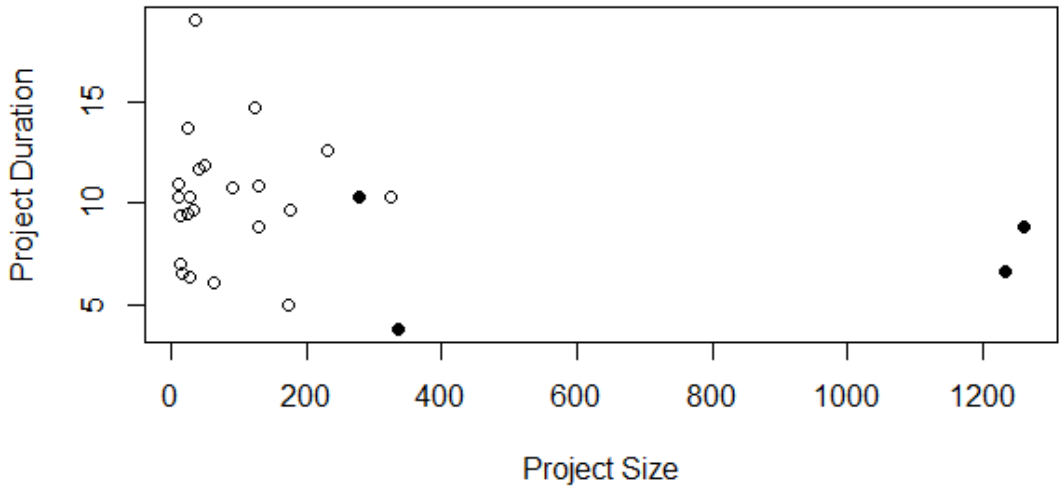
7.2.13 Plot of Size versus Cost

```
> plot(Repository$Size, Repository$Cost)
```



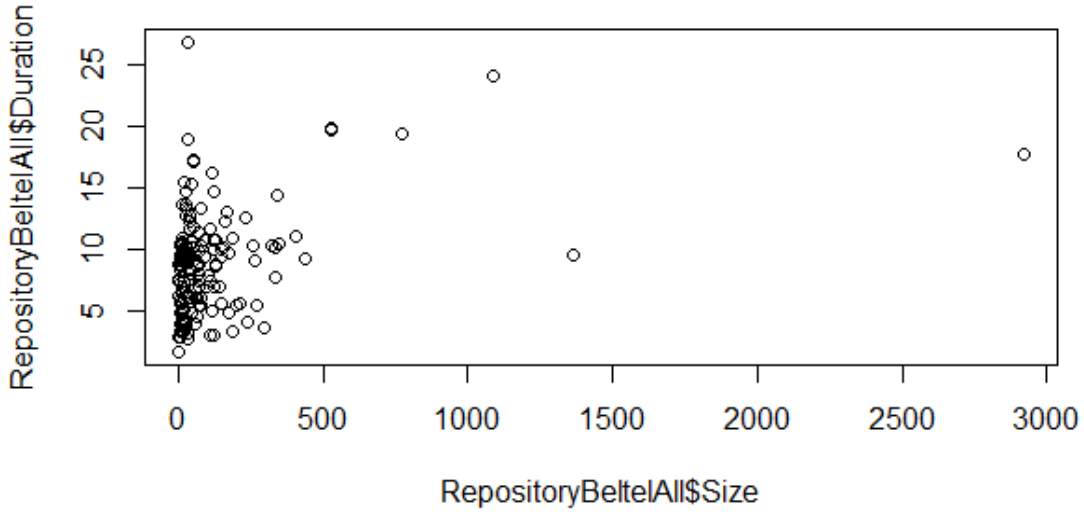
7.2.14 Plot of Duration versus Size

```
> plot(Repository$Size, Repository$Duration,
+      pch = c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,16,16,16,16),
+      xlab = "Project Size", ylab = "Project Duration")
```



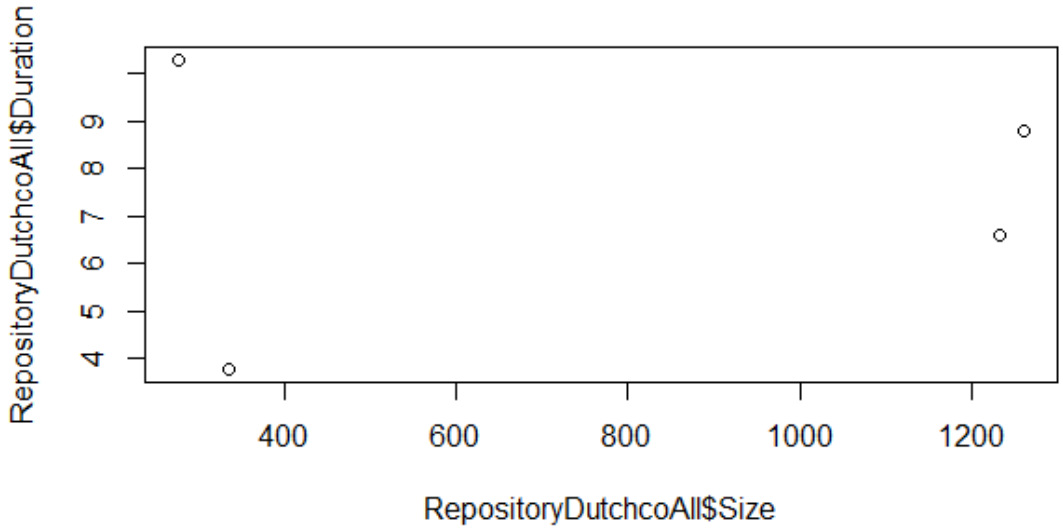
7.2.15 Plot of RepositoryBeltelAll$Size, RepositoryBeltelAll$Duration

```
> plot(RepositoryBeltelAll$Size, RepositoryBeltelAll$Duration)
```



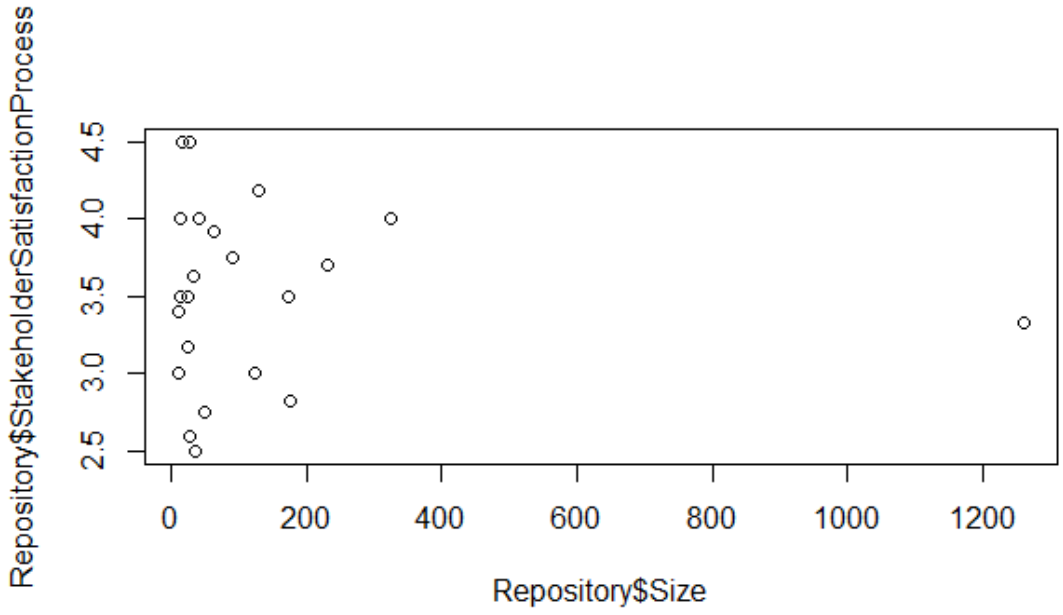
7.2.16 Plot of RepositoryDutchcoAll$Size, RepositoryDutchcoAll$Duration

```
> plot(RepositoryDutchcoAll$Size, RepositoryDutchcoAll$Duration)
```



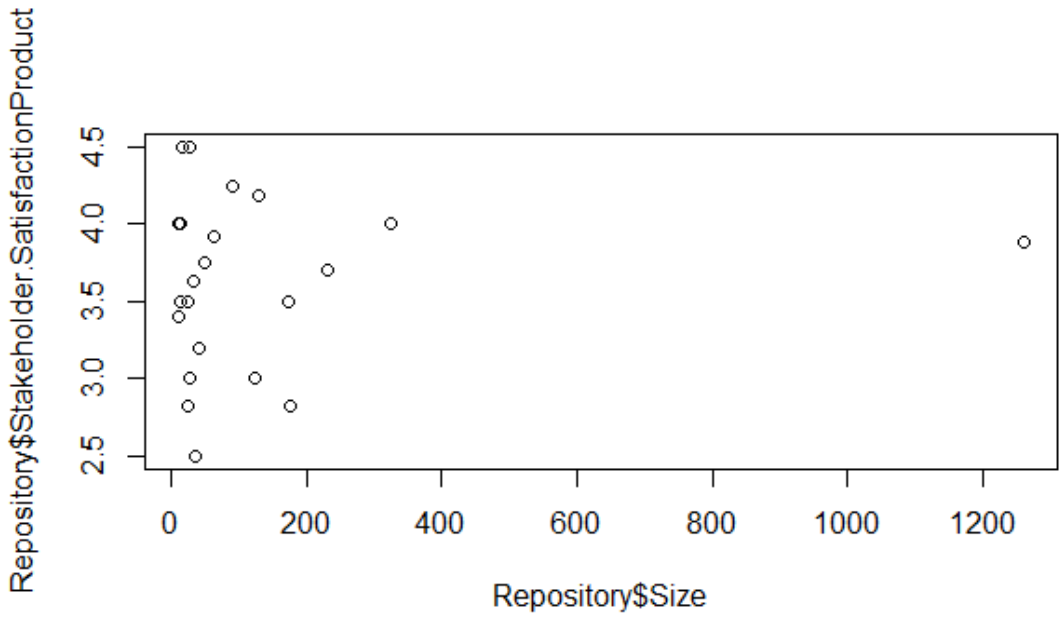
7.2.17 Plot of Repository$Size, Repository$StakeholderSatisfactionProcess

```
> plot(Repository$Size, Repository$StakeholderSatisfactionProcess)
```



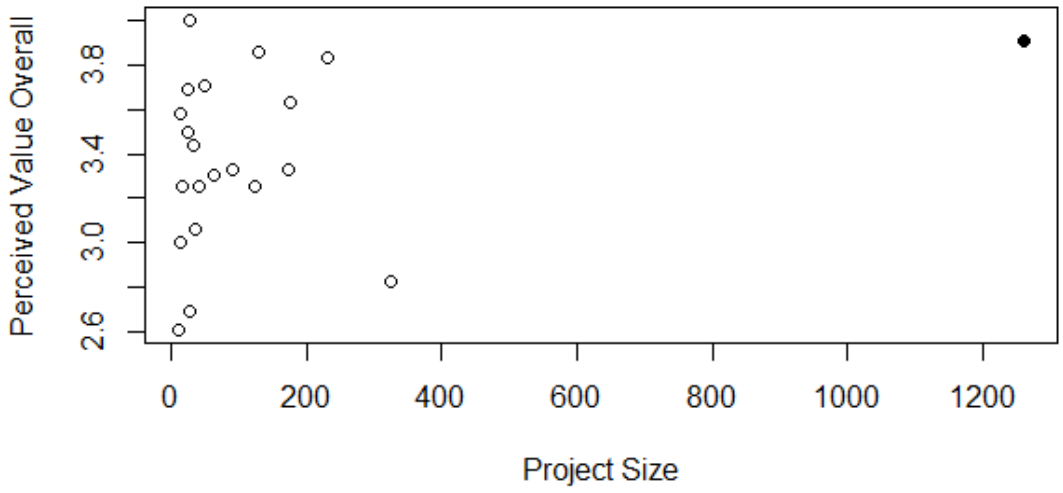
7.2.18 Plot of Repository$Size, Repository$Stakeholder.SatisfactionProduct

```
> plot(Repository$Size, Repository$Stakeholder.SatisfactionProduct)
```



7.2.19 Plot of Repository$Size, Repository$PerceivedValueOverall

```
> plot(Repository$Size, Repository$PerceivedValueOverall,
+      pch = c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,16,16,16,16),
+      xlab = "Project Size", ylab = "Perceived Value Overall")
```



7.2.20 Shapiro Wilk Normality test

```
> shapiro.test(Repository$Cost)

        Shapiro-Wilk normality test
```

```
data:  Repository$Cost
W = 0.60549, p-value = 3.423e-07

> shapiro.test(Repository$Size)

        Shapiro-Wilk normality test

data:  Repository$Size
W = 0.54617, p-value = 7.492e-08
```

TUD-SERG-2017-001
ISSN 1872-5392